The Impact of Acute Psychological Stress on Declarative and Working Memory Functioning

Robyn Human
HMNROB001

A minor dissertation submitted in partial fulfillment of the requirements for the award of the degree of MA in Psychological Research

Faculty of the Humanities
University of Cape Town
[2009/2010]

COMPULSORY DECLARATION

This work has not been previously submitted in whole, or in part, for the award of any degree. It is my own work. Each significant contribution to, and quotation in, this dissertation from the work, or works, of other people has been attributed, and has been cited and referenced.

Signature:_____ Date:_____

Table of Contents

List of Figures

List of Tables

Abstract

Previous research has shown that stress affects processing in many different memory systems. This study aimed to investigate the effects of acute psychosocial stress on declarative memory (DM) and working memory (WM) performance, and to explore whether sex differences exist under stress in these two memory systems. DM was assessed using cued recall and recognition of a verbal paired-associates list. WM was assessed using an *n*-back test with various difficulty levels. One hundred (42 males) undergraduate psychology students from the University of Cape Town were recruited. Phase of menstrual cycle and oral contraceptive use were controlled for in female participants. Participants took part in two sessions, 24 hours apart, each beginning after 16h00. Day 1 involved learning and immediate cued recall of the word pairs, and completing a practice *n*-back protocol. During Day 2, 45 participants were exposed to a psychosocial stressor and 41 were exposed to a relaxation period. Physiological and self-report measures of stress were taken at three intervals pre- and post-experimental manipulation. Participants then completed delayed cued recall and recognition tests for the previously-learned word pairs, and the full version of the *n*-back test. Data were analysed only for participants characterised as "cortisol responders" following the experimental manipulation. The final sample included 57 participants (30 males). With regard to DM, stress did not affect either delayed cued recall or recognition performance in either men or women. With regard to WM, stress negatively affected accuracy among men, but not women. These results are largely consistent with previous literature, but also elucidate a sex difference in working memory performance under stress (viz., while men's performance is negatively affected by stress, women show improved performance). The study provides important evidence for sex differences in WM performance under stress, and highlights several methodological issues that should be addressed in future studies.

Introduction

Effects of Stress on Functioning

The effects of stress on memory functioning have been documented in a number of different situations. One of the diagnostic criteria for Posttraumatic Stress Disorder (PTSD), which occurs in response to traumatic, stressful events such as rape or torture, is the "inability to recall an important aspect of the trauma" (American Psychiatric Association [APA], 1994, *p*. 428). However, such memory impairment does not only occur in the face of major, traumatic stressors. The negative effects of relatively minor stressors on memory functioning are visible in everyday life. Almost everyone has experienced the situation of studying diligently for an important exam, only to walk into the venue feeling stressed and anxious and to discover that when one looks at the questions on the paper, one cannot remember the answers. Thus, the experience of memory impairment as a result of everyday life stressors has practical implications for understanding how cognitive processes work. As a result, this subject has been widely researched for a number of years, although much remains to be discovered.

A stressor is an environmental event, or the anticipation of such an event, that results in one of two general stress responses. The first of these responses quickly results in major physiological manifestations that include sweating and an increasingly rapid heartbeat. The second response is a slower cognitive reaction involving the functioning of the hypothalamic-pituitary-adrenal (HPA) axis. The HPA axis response leads to the secretion of hormones involved in increasing the ability of an individual to handle a crisis situation. It is important to note that although these hormones are produced in great quantities at times of stress, they are also produced and sustained by the HPA axis under normal circumstances (Alderson & Novack, 2002; Kemeny, 2003; Kudielka & Kirschbaum, 2005; Sapolsky, 2004; Wolf, 2003).

Basic HPA axis functioning progresses along the following paths: First, the detection of a stressor causes the hypothalamus to release corticotrophin-releasing hormone into the anterior pituitary, which in turn releases adrenocorticotrophin into the blood. This release results in the synthesis of hormones (viz., glucocorticoids; cortisol in humans) in the adrenal cortex. From here these hormones move through the blood-brain barrier into the brain where they impact on a number of different areas, including two that play important roles in intact memory functioning: the hippocampus and prefrontal cortex (PFC; Alderson & Novack, 2002; Kemeny, 2003; Sapolsky, 2004; Wolf, 2003).

The hippocampus is critical for declarative memory (DM) functioning which involves the processing of specific information that is easily remembered and that can be expressed verbally (Squire, 1992). The PFC has been implicated in working memory (WM; Owen, 1997) which involves the short-term retention and handling of limited bits of information (Baddeley, 2001). Because these structures are involved in such vital cognitive functions, it is important to investigate the effects that stress may have on them and the processes to which they are critical. Thus, research in this area is vital in order to better understand human functioning.

Factors Affecting HPA Axis Responses to Stress

There are a number of factors that can affect responses to stress, and that may therefore moderate or mediate the impact of acute stress on memory functioning. These include the gender and age of participants, the amount of time spent under stress, the type of stressor used, the memory system being studied, and the time of day at which the research is conducted (Dickerson & Kemeny, 2004; Kudielka, Hellhammer, & Wüst, 2009; Kudielka & Kirschbaum, 2005; Kudielka, Schommer, Hellhammer, & Kirschbaum, 2004; Maheu, Collicut, Kornik, Moszkowski, & Lupien, 2005; Wolf, 2003). In this study, the effects of gender will be investigated directly, and effects of time of day will be controlled for explicitly.

Sex differences. Studies have shown that in response to an acute psychosocial stressor, men exhibit greater HPA axis responses (i.e., greater increases in cortisol levels) compared to women (Kirschbaum, Wüst, & Hellhammer, 1992; Uhart, Chong, Oswald, Lin, & Wand, 2006). However, in the Uhart et al. (2006) study, the women involved were all in the follicular phase of the menstrual cycle, while in the Kirschbaum et al. (1992) study, the female sample included women at all different stages of the menstrual cycle and also included women who were using oral contraceptives. These facts call into question the validity of the cortisol-related conclusions drawn from these studies as it has been shown in other research that the use of oral contraception and the phase of the menstrual cycle can influence stress-induced cortisol levels.

Specifically, Kirschbaum et al. (1995) found that women using oral contraception exhibited smaller increases in cortisol compared to men and to women not using oral contraception. In addition, Kirschbaum, Kudielka, Gaab, Schommer, and Hellhammer (1999) investigated the effects of the use of oral contraception and phase of menstrual cycle (namely the

follicular and luteal phases) on HPA axis functioning. Their results indicated that although under normal circumstances there are scarcely any noticeable differences between the two sexes in terms of HPA axis functioning, this is not the case in instances of stress: Women in the luteal phase were closely matched with men in terms of stress-induced salivary cortisol levels, although as with the Kirschbaum et al. (1992) and Uhart et al. (2006) studies, men did exhibit a slightly greater cortisol increase. However, perhaps more significantly, the levels for these two groups were higher than those for the group of women taking oral contraceptives and the group of women in the follicular phase of their cycle.[1]

Thus, it appears that sex differences in HPA axis functioning in reaction to stress do exist and should be investigated in studies involving memory. Additionally, the use of oral contraceptives and stage of menstrual cycle should also be controlled for in these studies, as these can also impact on stress-induced cortisol level increases. Unfortunately, studies controlling for the differential sex-based effects of stress on HPA axis activity are generally lacking.

Time of day. Cortisol has a circadian rhythm with levels peaking in the morning just after one awakes, and decreasing slowly over the course of the day, leading to the lowest levels in the late afternoon and evening (Dickerson & Kemeny, 2004; Kudielka et al., 2004; Lupien et al., 2002; Maheu et al., 2005). Thus, it has been suggested that studies using acute psychosocial stressors should take place in the late afternoon, after 16h00 if possible, when cortisol levels are at their lowest and most constant as this is when cortisol changes due to a stressor will be most easily identified (Dickerson & Kemeny, 2004; B. M. Kudielka, personal communication, June 5, 2008; Kudielka et al., 2009).

The results of some studies, such as Kudielka et al. (2004) and Maheu et al. (2005), appear to contradict this suggestion, however. These studies showed that although baseline salivary cortisol levels were higher in the morning than in the afternoon, there did not appear to be a great variance in cortisol level increase in response to an acute stressor between participants

[1] It is important to note that these sex differences in cortisol reactivity appear to only exist for salivary cortisol measurements, with studies generally showing that increases in plasma cortisol levels do not differ between men and women (Kelly, Tyrka, Anderson, Price & Carpenter, 2008; Kirschbaum et al., 1999; but see Uhart et al., 2006, for contrasting results).

run in the morning and those run in the afternoon. They did find, however, evidence that greater baseline levels (i.e., such as those generally recorded in the morning) may result in a slightly smaller stress response.

Possibly as a result of the above contradictions, studies in this area have not been consistent in the times at which they have been run, making comparisons between them problematic. For instance, as already stated, it has been suggested that studies using acute stressors should preferably take place after 16h00 (B. M. Kudielka, personal communication, June 5, 2008). However, very few researchers conduct their studies at that time, with most, including Maheu et al. (2005), being conducted earlier in the day. Although the three afternoon studies included in Kudielka et al.'s (2004) review (Kirschbaum et al., 1999; Kudielka, Schmidt-Reinwald, Hellhammer, & Kirschbaum, 1999; Kudielka, Schmidt-Reinwald, Hellhammer, Shürmeyer, & Kirschbaum, 2000) all began between 15h00 and 16h00, it appears that the reviewers excluded the data from 12 young male participants from the Kudielka et al. (2000) study. They do not appear to provide a reason for this exclusion, although the salivary cortisol responses of these young males, as reported in the original article, indicate that they may have had an effect on the overall results of the review.

In addition, of the studies mentioned above, only Maheu et al. (2005) investigated the effects of stress on memory, while the rest only investigated physiological responses to an acute psychosocial stressor in different populations (i.e., they had no cognitive component). Therefore, it is difficult to know what impact on memory performance the results from the other studies might have had. However, contrary to what predictions the above information appear to suggest (viz., that because cortisol increases are comparable in the morning and the afternoon, it would seem that there should not be a significant difference in memory performance under conditions of stress between these two times, or that if a slightly larger stress response is seen during the afternoon then it would seem that this should be when decreased memory performance would be seen), Maheu et al. (2005) found that while memory for emotional stimuli was impaired after exposure to an acute stressor in the morning, this was not the case in the afternoon. They used a variation (outlined in de Kloet , Oitzl, & Joëls, 1999) of the inverted-U hypothesis (originally outlined by Yerkes & Dodson, 1908) to explain this finding, arguing that because baseline cortisol levels are higher in the morning, corticosteroid receptors in the brain are already experiencing high levels of activation. Therefore, increases in cortisol levels as a result of a

stressor at this time of day will lead to the corticosteroid receptors becoming fully, or almost fully, activated, leading to memory impairment. However, because of the lower baseline levels of cortisol in the afternoon, corticosteroid receptors are experiencing lower general levels of activation than in the morning. Therefore, increases in cortisol levels as a result of a stressor at this time of day will lead to a lower, more optimal level of post-stressor corticosteroid receptor activation than in the morning, resulting in no impairment in memory.

Nonetheless, findings in this field are still somewhat inconclusive and more studies specifically controlling for and investigating time of day-related effects of stress on cognitive functioning are needed.

Effects of Stress on Memory

Research studies investigating the effects of stress on memory achieve an increase in participants' cortisol levels using several different methods. These methods can include using naturally occurring stressors during which to conduct their tests, such as examination periods (e.g., Vedhara, Hyde, Gilchrist, Tytherleigh & Plummer, 2000) and the cold-pressor test, which involves placing participants' hands in very cold water (e.g., Porcelli et al., 2008). However, the two methods which have proved to be most popular are (a) through the direct administration of cortisol (e.g., orally in the form of cortisone) to participants (e.g., de Quervain et al., 2003; de Quervain, Roozendaal, Nitsch, McGaugh, & Hock, 2000; Kirschbaum, Wolf, May, Wippich, & Hellhammer, 1996, Study 2; Lupien, Gillin, & Hauger, 1999; Lupien et al., 2002, Study 2; Tops et al., 2003; Wolf, Convit, et al., 2001), and (b) to expose participants to a psychosocial stressor, such as the Trier Social Stress Test (TSST; e.g., Elzinga & Roelofs, 2005; Kirschbaum et al., 1996, Study 1; Kuhlmann, Piel, & Wolf, 2005; Luethi, Meier & Sandi, 2009; Nater et al., 2007; Oei, Everaerd, Elzinga, Van Well, & Bermond, 2006; Schoofs, Preuβ, & Wolf, 2008; Wolf, Schommer, Hellhammer, McEwen, & Kirschbaum, 2001). The studies reviewed below each used one of these two forms of experimental manipulation of cortisol levels.

Declarative memory. Although a number of studies have investigated the effects of increased cortisol levels on DM through the use of free and cued recall tasks, findings have been largely inconclusive. More specifically, results appear to differ according to which task is being

investigated and at what point in the memory process cortisol levels are increased. Unfortunately, however, even results within these specifications are not consistent.

With regard to the free recall of verbal information, several studies have shown that delayed recall is (a) impaired when cortisol levels are increased just before the retrieval phase of memory processing, but is (b) generally not affected when they are raised at other points (e.g., before the encoding phase; de Quervain et al., 2000; Kuhlmann et al., 2005; Luethi et al., 2009; Nater et al., 2007; Wolf, Convit, et al., 2001; Wolf, Schommer, et al., 2001). In terms of immediate free recall, it has generally been shown that this remains unimpaired whether cortisol levels are raised at either the encoding or the retrieval stages of the memory process (de Quervain et al., 2000; Nater et al., 2007; Wolf, Convit, et al., 2001).

However, both Nater et al. (2007) and Wolf, Schommer, et al. (2001) found that with closer analysis of their results, correlations between cortisol levels and memory performance could be observed where the bigger between-groups analyses showed no statistically significant results. Specifically, the latter study found some evidence for a negative correlation between cortisol responses to stress experienced at the encoding phase and delayed free recall performance, while the former study found a similar correlation for immediate free recall performance. In contrast with the other studies but in line with these latter results, Tops et al. (2003) found that immediate free recall was affected by increased cortisol levels at the encoding stage. In summary, it appears that with regards to free recall, although there are some inconsistencies, stress experienced at the point of encoding does not generally affect performance on delayed or immediate free recall tasks, but stress experienced at the point of retrieval does tend to affect performance on delayed recall of these tasks.

With regard to the delayed cued recall of verbal information, while two studies indicated that higher cortisol levels prior to the encoding of information resulted in a decreased performance on such tasks (Kirschbaum et al., 1996), Lupien et al. (1999) found that elevated cortisol levels at this phase did not affect this type of memory. Similarly to the results for the effects of stress on delayed free recall, de Quervain et al. (2003) found that 24-delayed cued recall was impaired by increased cortisol levels at the retrieval stage. However, Kuhlmann et al. (2005) found that delayed cued recall remained unaffected by stress experienced just before memory retrieval. Thus, results for this memory type appear to be far more inconclusive than those for the free recall of verbal information.

In comparison to verbal recall memory, the effects of stress on the recognition of previously learned verbal information have not been as extensively examined. Studies have generally found that overall recognition performance is unaffected by increased cortisol levels regardless of whether cortisol increases are induced at the encoding (de Quervain et al., 2000; Luethi et al., 2009; Nater et al., 2007; Tops et al., 2003) or retrieval (de Quervain et al., 2000, 2003; Lupien et al., 2002, Study 2) stage.

There is, however, some evidence that recognition for particular types of stimuli may in fact be affected by stress. Tops et al. (2003) found that increased cortisol levels resulted in significantly decreased recognition of neutral words but, generally resulted in better recognition of unpleasant words. In addition, Lupien et al. (2002, Study 2) found that increased cortisol levels led to improved reaction times when making decisions on a recognition task. Although this is a measurement that most studies in this field do not include, Tops et al. (2003) did include it but found no differences in reaction time data for their groups.

Although recognition performance under stress does not appear to have been extensively studied, Lupien et al. (1999) suggested that including a recognition task in conjunction with a recall task could be advantageous. This inclusion would allow one to compare the effects of stress on these two memory forms and to better understand the effects of stress on recall performance, especially if stress is induced at the level of encoding: If the stressor were to affect the memory process at the point of encoding, then both recall and recognition performance should be impaired, because the words would not have been properly encoded. But, if the stressor were to affect the memory process at the point of retrieval only, then while recall would be affected, recognition should be (relatively) intact because the words would have been properly encoded and the cues provided by the recognition task would facilitate remembering resulting in the material being likely to be recognized if not recalled. Thus, although most studies show that stress does not impact recognition performance, regardless of the stage of the memory process at which cortisol levels are increased, the effects of stress on recognition should be further investigated and should include analyses of participants' reaction times during the recognition tasks.

Sex differences in DM studies. On most intelligence tests, men and women perform differentially on different subtests, indicating a sex-based variation in different cognitive abilities, rather than an overall intellectual superiority of one of the sexes (Halpern & LaMay,

2000). For decades it has generally been found that men are better at spatially-based tasks and that women are better at verbally-based tasks (Maccoby & Jacklin; 1974; Weiss, Kemmier, Deisenhammer, Fleischhacker & Delazer, 2003).

In terms of DM, a number of studies over the past few decades have found that women tend to perform better than men. For example, Bleecker, Bolla-Willson, Agnew, and Meyers (1988) found that across different ages of adulthood, females performed better on immediate verbal recall than did males, although they found no differences in recognition ability between the sexes. More recently, Maitland, Herlitz, Nyberg, Bäckman, and Nilsson (2004) found that females tended to perform significantly better than males on a number of memory tasks, both episodic and semantic in nature, including recognition, recall, and verbal fluency tasks. They also found that these differences tended to decrease with an increase in age and that these decreases were especially noticeable with regard to recall performance, while the differences in recognition performance tended to remain more stable.

Although these sex differences in DM performance under normal circumstances are well documented, studies adequately investigating sex differences in the effects of stress on DM performance are very scarce. More than half of the studies reviewed above (de Quervain et al., 2003; Kirschbaum et al., 1996, Study 2; Kuhlmann et al., 2005; Luethi et al., 2009; Lupien et al., 1999, 2002, Study 2; Nater et al., 2007; Tops et al., 2003; Wolf, Convit, et al., 2001) only used male participants and so could not investigate the question of whether sex differences in DM performance exist under conditions of increased cortisol levels. Furthermore, although both de Quervain et al. (2000) and Kirschbaum et al. (1996, Study 1) included both male and female participants in their samples, neither controlled for the use of oral contraceptives or stage of the menstrual cycle. As noted above, these controls are thought to be important in order to make comparisons between the sexes and to therefore draw accurate conclusions about any differences in performance under stress. In addition, the Kirschbaum et al. (1996) study did not have a control group. Thus, although this study found their female participants tended to recall a greater number of words than their male participants (who experienced the greater cortisol increases than the female participants as would be expected), this result needs to be interpreted with caution, especially as it was not statistically significant. Nonetheless, that study's data does suggest that there may a difference in the DM abilities of the sexes under conditions of stress,

although adequate control measures need to be included in future studies to ascertain the true extent and nature of this difference.

Only Wolf, Schommer, et al. (2001) attempted to control for the abovementioned variables. They excluded females who were using oral contraceptives or who were not in the late luteal phase, defined in their study as being days 21 to 25, of their menstrual cycles. Although, as noted above, they did not find an overall effect of stress on DM performance, they did find that there was a negative correlation between free recall test performance and cortisol responses to stress in male participants. This finding suggests that cortisol does affect DM performance, but only in males. In addition, the authors state that this observed difference between the sexes cannot be accounted for by differential increases in cortisol levels as these were controlled for through their exclusion criteria for female participants. Thus, the results of this study, in addition to the results of the Kirschbaum et al. (1996) study, imply that there may be sex differences in the effects of stress on DM performance. This finding has not been replicated to this point.

Time of day in DM studies. As previously discussed, the time of day at which studies investigating the effects of stress on memory are conducted could have an impact on the results. Thus, this factor needs to be taken into consideration when interpreting the results of such studies. Of the studies using psychosocial stressors reviewed thus far, two of them (Kuhlmann et al., 2005; Wolf, Schommer, et al., 2001) were conducted between 10h00 and about 12h30.

Of the studies that did take place later in the afternoon (Kirschbaum et al., 1996, Study 1; Luethi et al., 2009; Nater et al., 2007), Luethi et al. (2009) only ran a third of their participants after 16h00, and just under a third before 14h00. They state that as a result of these time variations, baseline cortisol levels did differ between participants but this did not affect the cortisol increases in response to the stressor. Clearly, however, when one takes into account the circadian rhythm of cortisol secretion, one cannot be sure that these variations in time did not impact on their memory results. Nater et al. (2007) ran their 90-minute experimental procedures between 14h00 and 18h00. Therefore, if participants began the study at 14h00, they would have been finished by about 15h30. The fact that some participants would have been tested before 16h00 and some would have been tested after 16h00 means that there could have been differences in their baseline cortisol levels, which could have affected their HPA axis responses to the stressor and thus impacted upon the results. In addition, although Kirschbaum et al. (1996,

Study 1, p. 1476) state that their study took place "during late afternoon hours", they do not give a specific time.

Although there are too few studies here to draw any definitive conclusions, it is interesting to note that, contrary to other research, on average these three afternoon studies showed baseline cortisol levels similar to the two studies conducted in the morning (afternoon: *M* = ± 9.5 nmol/l; morning: *M* = ± 8.5 nmol/l).However, as suggested by Kudielka et al. (2004) and Maheu et al. (2005) might occur, the afternoon studies (which had slightly larger basal cortisol levels) showed slightly larger average increases following the stressor (afternoon: *M* = ± 11.5 nmol/l; morning: *M* = ± 8 nmol/l) although it does not appear that the results of the afternoon and morning studies show any distinctive differences in their results. However, it is clear that studies conducted exclusively in the later afternoon in this area are lacking, and greater time of day controls are needed in order to further clarify existing results.

Working memory. Studies on the effects of increased stress and cortisol levels on WM generally use one of three main measures: digit span tests (using both forwards and backwards conditions), the *n*-back test, and the 'Sternberg paradigm'. The digit span forwards task involves participants being read a random set of numbers and then having to repeat them back in the same order. The digit span backwards task involves participants being read a random set of numbers and then having to repeat them in the reverse order to the presentation. The task increases in difficulty as the number sets increase in length (Psychological Corporation, 1998). In the *n*-back task, participants are presented with a series of letters or digits and are required to indicate whether the one currently being presented matches the letter or digit from *n* letters or digits before. Thus, in a 1-back task, participants would need to indicate if the current letter or digit is the same as the letter or digit presented before it, while in a 2-back task, participants would need to indicate if the current letter or digit is the same as the letter or digit presented two letters or digits previously (Schoofs et al., 2008). Generally, performance on this task decreases with an increase in task difficulty (Speck et al., 2000). In the original Sternberg paradigm, participants were presented with a set of digits that they were required to remember. After a brief delay, they were shown another digit and were required to indicate if it matched any in the first set of digits. The first set of digits varied in length, thereby varying the task difficulty (Sternberg, 1966). In more recent studies, however, this paradigm involves participants being presented with between

one and four letters to remember and then after a delay are presented with a second set of between one and four letters and are required to indicate whether any of the first set of letters is present in the second set of letters. Again, the task varies in difficulty according to how many letters need to be remembered and how many are presented in the second set (Lupien et al., 1999; Oei et al., 2006).

With regard to digit span tests, there is some debate about their efficacy as a measure of WM. Many in the field argue that the forwards digit span task is not a true test of WM, because it does not require the manipulation of information (Jarrold & Towse, 2006; Lynn & Irwing, 2008). Furthermore, Schoofs et al. (2008) argue that digit span tests do not include measurement of reaction times, a factor they believe to be important in investigating WM. They also state that digit span tests are shorter than either of the other two tasks usually used to measure WM, with each difficulty level typically being tested by only two trials. The combination of these factors suggests that digit span tests may not be as sensitive a measure of WM as, for instance, the Sternberg paradigm or the *n*-back task.

Although there is some debate as to whether or not the lower level of the *n*-back task (i.e., the 1-back condition) really is a test of WM, the *n*-back task is generally considered to be a true test of WM and is often used for research purposes (Jarrold & Towes, 2006; Schoofs et al., 2008). Neuroimaging studies have shown that performance on the *n*-back task results in PFC activation, and that brain activation increases as the task difficulty level increases. This latter fact provides further evidence that this task does test WM ability, especially at higher difficulty conditions (Braver et al., 1997; Cohen et al., 1997; Jarrold & Towse; 2006; Speck et al., 2000).

Despite these debates as to which WM paradigm should be used for research, studies on the effects of stress on WM have reached more conclusive results than have studies on the effects of stress on DM. Luethi et al. (2009), Lupien et al. (1999) and Wolf, Convit, et al. (2001) investigated the effects of increased cortisol levels on both DM and WM. The first study used a reading span task as their measure of WM, the second study used the contemporary Sternberg paradigm, and the third study used the forward and backward digit span. In contrast to their findings regarding DM, basic results for all three studies showed that increased cortisol levels resulted in a definite impairment of WM. However, whereas Luethi et al. (2009) and Lupien et al. (1999) explicitly and intentionally compared the effects of increased cortisol levels on WM and DM, Wolf, Convit, et al. (2001) did not. The latter authors admitted that they "did not

anticipate that cortisol would have much of an effect on frontal function" and as such, "Digit Span was being used as a control task" (p. 1008). They go on to state that other, more refined tests of WM should be used to investigate these effects more closely.

Most of the recent studies in this field, with the exception of Kuhlmann et al. (2005), have supported these findings that increased cortisol levels negatively affect WM performance. For instance, Elzinga and Roelofs (2005) used a digit span task and found increased cortisol levels to be detrimental to WM functioning, specifically with regard to the forward condition. Oei et al. (2006), using the contemporary Sternberg paradigm, found that at greater WM loads, increased cortisol levels resulted in both greater WM impairments and slower reaction times.

Schoofs et al. (2008) found further evidence for the impairment of WM under stress. They used the *n*-back test (with two difficulty levels, a 2-back and a 3-back condition) to assess WM. Compared to unstressed participants, their stressed participants showed a significant impairment in WM performance (in terms of both correct responses given and reaction times) across both difficulty levels. Interestingly, however, the authors noted that the effects of stress on WM ability decreased during the duration of the task. This finding echoed the pattern detected by Elzinga and Roelofs (2005), who observed that for both stressed and non-stressed participants, WM improved as the task became more familiar. Schoofs et al. (2008) speculated that this pattern of performance may have arisen because WM tasks are more likely to be affected by stress when they are novel to the participant than when they are practiced.

Schoofs et al. (2008) also noted that it may be useful to include *n*-back conditions that range from easy (e.g., 1-back) to more difficult (e.g., 3-back) in order to identify at what difficulty level WM is affected. This suggestion is based on Sliwinski, Smyth, Hofer, and Stawski's (2006) finding that although impaired performance due to stress could be seen on more difficult *n*-back conditions (e.g., 2-back), performance remained intact on easier conditions (e.g., 1-back).

Thus, it is clear that although empirical findings generally support the fact that WM is impaired by stress, especially when more pressure is put on the WM systems (e.g., at greater processing loads), there is great inconsistency in the measures being used. As previously discussed, some of these measures may be more efficacious than others and because a number of studies use the digit span task as their measure, they may not be truly measuring WM in its entirety. Thus, findings across studies may not be comparable and greater control over the types

of measures used needs to be exerted in future studies to ensure that what is being measured truly is WM.

Sex differences in WM studies. In terms of WM performance under non-stressful conditions, results appear to be mixed. Lynn and Irwing (2008) found that men performed slightly better than women in digit span tasks. This finding, coupled with the rest of their results, led them to conclude that men have a greater WM capacity than women. In contrast though, Speck et al. (2000) found that for *n*-back tasks, regardless of difficulty level, female participants showed better performance in terms of accuracy, although they did perform slightly slower than the male participants. Thus, it appears that sex differences in WM may exist under normal circumstances.

As is the case for DM, sex differences in the effects of stress on WM have not been sufficiently investigated. Again, most of the studies in this area have included only male participants (Luethi et al., 2009; Lupien et al., 1999; Kuhlmann et al., 2005; Oei et al., 2006; Schoofs et al., 2008; Wolf , Convit, et al., 2001). The only WM study reviewed here that used both male and female participants was that conducted by Elzinga and Roelofs (2005), although their control for both the use of oral contraceptives and the stage of the menstrual cycle (day 21-25) was only done in an attempt to reduce any differences in the results that may have arisen due to gender. Although their data showed little difference between the sexes in terms of WM performance, the authors make the observation that their sample size (stressed group: $n = 9$) was too small to draw any definite conclusions.

It is clear that even if studies do not specifically investigate sex differences in the effects of stress on WM, studies that (a) include both sexes, and (b) controlling for the relevant extraneous variables are necessary to give a more well-rounded account of this area. However, clearly the investigation of actual differences in WM performance between the sexes is also necessary for a complete picture to be achieved.

Time of day in WM studies. Of the WM studies reviewed here that used a psychosocial stressor, most of them were conducted between 09h00 and about 12h30 (Elzinga & Roelofs, 2005; Kuhlmann et al., 2005; Oei et al., 2006; Schoofs et al., 2008). Only Luethi et al. (2009) conducted their study later in the day, but as already pointed out, only a third of their participants underwent the procedure after 16h00. Despite this fact, again the average cortisol levels of the morning studies were similar to those of Luethi et al. (2009; morning: $M = \pm 11.5$ nmol/l;

afternoon = 10 nmol/l) and again the average morning cortisol increases in response to the stressor were smaller than in Luethi et al. (2009; morning: *M* = ± 8 nmol/l; afternoon = 13.2 nmol/l). Also, there are no differences in results between the one afternoon study and most of the morning studies, although the fact that some of the Luethi et al. (2009) participants were run earlier in the day than 16h00 might account for the similarity in results.

Therefore, it is once again clear that studies conducted exclusively in the late afternoon are lacking in this area. As noted before, studies using psychosocial stressors that are conducted in the late afternoon could provide a different pattern of cortisol reactivity (and consequently a different pattern of cognitive performance) compared to studies that are conducted earlier in the day. Thus, if such studies were to be performed, they may provide some interesting comparisons to the currently existing data, including the possibility of different patterns of cognitive results.

Rationale for Research

The literature reviewed above suggests that there is largely conflicting evidence regarding the effects of stress on different memory processes. In addition, it is clear that there are a number of confounding variables that need to be considered when designing studies in this area. My study, which looked at the effects of an acute psychosocial stressor on verbal DM and WM, addressed many of the methodological limitations present in previously conducted studies and improved and extended these studies in several ways, including the following:

1) Although studies investigating the effects of stress on free recall DM show relatively consistent results, studies investigating the effects of stress on cued recall DM appear to be scarcer and more inconsistent in their findings. In addition, many of the previously published studies in this area artificially increased cortisol levels (i.e., through the administration of, for example, hydrocortisone) and therefore may not represent naturally occurring consequences of stress on memory performance. Thus, this study aimed to investigate the effects of an acute psychosocial stressor on cued recall DM, using a task similar to that described by Lupien et al. (1999). However, whereas those researchers investigated how increased cortisol levels at the encoding phase affected 15-minute delayed cued recall, this study investigated how stress experienced at the retrieval stage would affect 24-hour delayed cued recall. The longer delay in the current study meant participants would have time to fully consolidate the word pairs before cued recall was initiated (Luethi et al., 2009), thereby truly testing long-term memory performance.
2) In addition to the cued recall task, this study also included a recognition task as a test of DM. This inclusion aimed to investigate Lupien et al. (1999)'s hypothesis that stress experienced at the retrieval stage would not affect recognition performance. The test of recognition used here also included a reaction time measure, which few previous studies have done, but which has been found to be potentially affected by stress. In addition, instead of looking solely at the recognition hit rates or separately at the recognition hit and false alarm scores of the participants as is done in some other studies (e.g., de Quervain et al., 2000, 2003), I used a d-prime analysis on the data in order to investigate how well the participants discriminated between the previously-learnt and newly-presented word pairs.

3) The impact of stress on memory tends to be investigated in terms of effects on DM performance more often than in terms of effects on WM performance (Schoofs et al., 2008), and therefore the latter area has not been as thoroughly researched as the former. The current study aimed to investigate the effects of stress on WM using a more efficacious measure than the digit span test, which has been used in a number of previous studies in this field. I followed Schoofs et al. (2008) in using the *n*-back paradigm to measure WM. Additionally, I addressed these latter authors' concerns about their limited use of this task by including a very easy *n*-back condition alongside a more difficult one.
4) Apart from the studies by de Quervain et al. (2000, 2003) and Kuhlmann et al. (2005), which were conducted over 2 days, the majority of other studies in this field have been conducted in 1 day, thereby potentially not allowing enough time for the consolidation of the learnt material. The current study was conducted over 2 days, allowing for 24-hour delayed cued recall to be tested. This design meant that I could study both between-group and within-group effects of stress on delayed cued recall DM, thereby extending the literature in potentially interesting ways. This design also allowed me to ensure that all participants understood the requirements and nature of the WM task before undergoing the experimental manipulation.
5) Unlike the majority of previously conducted studies, my study looked at sex differences in the effects of stress on verbal DM and WM while controlling for the use of oral contraceptives and menstrual cycle stage. Accordingly, this study aimed to not only look at how stress affects these functions, but to also elucidate more fully whether true sex differences in memory functioning exist under stressful circumstances.
6) This study aimed to employ a larger sample size than in the majority of previously conducted studies. The average sample of size for the DM studies reviewed above is 37 people, while the average sample size for the WM studies reviewed above is 39 people. I aimed to have a sample of over 50 participants, thereby allowing for greater statistical power and generalizability of results than previously published studies.
7) Unlike the majority of previously published studies, this study was conducted between 16h00 and 20h00 in order to control for time of day effects. This late afternoon phase was chosen (a) in order to avoid the morning peak in cortisol levels, and (b) because it has been suggested that it is easiest to obtain a cortisol increase from a psychosocial stressor

at this time (B. M. Kudielka, personal communication, June 5, 2008). In addition, conducting the study at this time means that baseline cortisol levels should be at a similar point in the cortisol diurnal cycle for all participants, and therefore avoids the confounding factor that may have been present in studies such as that of Luethi et al. (2009).

Given the overall rationale and the specific aims of the current study, my hypotheses were that:

1) Twenty-four hour delayed cued recall DM would be more impacted by stress than 24-hour delayed recognition, which would be unaffected by stress.
2) WM would be more negatively impacted by stress at higher levels of difficulty on the *n*-back task than at lower levels of difficulty.
3) Based on the findings regarding sex differences in verbal tasks under normal circumstances, I predicted that female participants in the stressed group would perform better than male participants in the stressed group on both the cued recall and recognition DM tasks.
4) Based on the findings regarding sex differences on the *n*-back task under normal circumstances, I predicted that (a) female participants in the stressed group would perform more accurately on the WM task than the male participants in the stressed group, but that (b) the female participants' performance on this task would be slower than the male participants' performance.

Design and Methods

Design and Setting

This study involved a 2 x 2 factorial design. The first independent variable was the sex of the participants (male or female). The second independent variable was the psychological state (stressed or relaxed) of the participants when undergoing memory testing on Day 2. Dependent variables were derived from participants' performances on both a WM measure (an *n*-back test with two main difficulty levels) and two verbal DM measures (cued recall and recognition scores on a verbal paired associates test).

Following de Quervain et al. (2003) and Kuhlmann et al. (2005), the study took place over 2 consecutive days for each participant. Day 1 consisted of (a) a learning phase for the verbal paired-associates (VPA) wordlist, and (b) a shortened version of the *n*-back test. The latter was included to ensure that all the participants understood how the WM task worked before undergoing the experimental manipulation on Day 2. On Day 2, the participants were pseudorandomly divided by sex into either the stress condition (Stress group) or the control condition (Relax group) and underwent the group-appropriate experimental manipulation before being administered the DM and WM testing phases.

In order to control for cortisol's circadian cycle and to elicit as strong an HPA axis response to the stressor as possible, the study was conducted between 16h00 and 20h00 (B. M. Kudielka, personal communication, June 5, 2008).

The study took place in two venues in the Department of Psychology at the University of Cape Town (UCT). The first was a computer laboratory where the memory testing, physiological measures, and questionnaire completion took place, and the second was the room where the participants underwent the experimental manipulation (the TSST or the relaxation period).

Participants

One hundred volunteer students (42 males), between the ages of 18 and 27 years, were enrolled in the study. Two additional participants were excluded immediately upon arrival for the study because they were females but had signed up by mistake on the male participant sign-up sheets and did not meet inclusion criteria. The participants were recruited from UCT undergraduate psychology classes by means of the Student Research and Participation Project (SRPP) as part of

their ‘duly performed’ certificate requirements. None of the participants were on any forms of steroid-based medication.

Because I sought to control for the use of oral contraception and phase of menstrual cycle in the female participants, sign-up procedures differed for male and female participants. Most of the male participants were recruited through sign-up sheets placed on the SRPP notice board on which they supplied their name, student number, email address, and contact number. Some of the male participants were recruited via an announcement posted on the first-year psychology course’s website or via information slips handed out during registration for the first-year psychology course. Students who were interested were asked to contact the researchers via email with the abovementioned information.

Female participants were recruited via either a notice on the SRPP notice board, an announcement placed on the first-year psychology course’s website, or information slips handed out during registration for the first-year psychology course. All females wishing to participate in the study and who were not on any form of oral or hormonal contraception were asked to contact the researchers via email with their name, student number, and contact number. This procedure ensured that the female participants’ privacy regarding their non-use of oral contraceptives was maintained. Once an email had been received, potential participants were then asked whether they had a regular menstrual cycle, and if so to indicate on what date they expected to begin their next period. I aimed to test all female participants in the late luteal phase of their menstrual cycles. This phase ends on the day preceding the start of the menses phase of the cycle and was defined as (approximately) the 6 days preceding this day (Bischof, 2003). Participants were asked to indicate on what date they expected to begin their next period and appointment times were then arranged for within the 6 days preceding that date. All female participants were asked to contact the experimenter on the first day of their next period for post-experimental verification of the phase of the menstrual cycle in which they were tested.

It must be noted that the definition of ‘late luteal phase’ used in this study was not identical to that used in other literature in this field. This phase is usually defined as days 21 – 25 of the menstrual cycle (e.g., Elzinga & Roelofs, 2005; Wolf, Schommer, et al., 2001) and Kirschbaum et al. (1999) defined the ‘luteal phase’ used in their study in those terms. However, I felt that my definition was more accurate for the following reasons:

a) Menstrual cycles vary in length from woman to woman with the average range varying from study to study. For women of the approximate age range of the female participants in my study, cycle length generally appears to be somewhere between 25and 31 days on average (see Table 1 for exact study details). Thus, it is clear that the day of the menstrual cycle on which the late luteal phase begins will vary from woman to woman , depending on the length of her cycle, and for some women will not fall between days 21 and 25 at all (Bischof, 2003).

Table 1
Between-Individual Variations in Menstrual Cycles and Menstrual Cycle Phases Across Three Studies.

Study	*n*	Age	Menstrual Cycle Length	Follicular Phase Length	Luteal Phase Length
Bischof (2003)	20	20-35	28 – 35	15.4 ± 2.5	13.6 ± 1.2
Cole et al. (2009)	167	18-36	20 – 34 (27.7 ± 2.4)	14.7 ± 2.4	13.2 ± 2.0
Sherman & Korenman (1975)	10	18-30	(30.00)	16.9 ± 3.7	12.9 ± 1.8

Note. Age range is given in years. Menstrual cycle, follicular and luteal phase lengths are given in days. Non-parenthetical menstrual cycle length values refer to range. Parentheses indicate mean ± standard deviation.

b) The length of an individual woman's menstrual cycle may vary from month to month within the same woman, with variations anywhere between 0.8 and 6.2 days (M = 2.8 ± 1.6 days; Cole, Ladner, & Byrn, 2009). This variation occurs because neither the follicular phase (from the start of menses until ovulation) nor the luteal phase (from the day of ovulation until the start of menses) are always stable. However, as can be seen from Tables 1 (between-women variations) and 2 (within-women variations), the luteal phase is generally more stable than the follicular phase both between- and within-women (Bischof, 2003; Cole et al., 2009). Thus, although the luteal phase is fairly constant within- and between-women, and the late luteal phase always ends on the day preceding the start of the menses phase of the cycle, the day on which the late luteal phase begins may vary from month to month. As a result, counting forward from the start of a woman's last period could be problematic if the next cycle (specifically the follicular

phase) is longer or shorter than expected. In addition, Cole et al. (2009) describe the fact that individuals who may be infertile or have problems with their menstrual cycles (both factors that individuals may not be aware of) may greatly increase the within-individual variation in menstrual cycle, making it even more difficult to predict the late-luteal phase of the cycle by working forwards.

Table 2
Within-Individual Variations in Menstrual Cycle Phases.

Study	*n*	Age	Follicular Phase	Luteal Phase
Cole et al. (2009)	167	18-36	0.5 – 18 (3.9 ± 3.7)	0.2 – 15 (2.6 ± 3.1)

Note. Follicular and luteal phase lengths are given in days. Non-parenthetical values refer to range. Parentheses indicate mean ± standard deviation.

Therefore, by counting backwards from the date on which the participants expected to start their next period, I could control for different cycle lengths and also, I hoped, control for variations in the length of the follicular phase in these participants' cycles.

Once the female participants' appointments had been booked based on their menstrual cycle calculations documented above, male participants were contacted via email and dates for their appointments were arranged around those of the already-booked female participants. Data collection occurred over a 7-month period and so the potential female participants continued to inform the researcher of when they expected to begin their following period until they could be accommodated.

Participants were pseudorandomly assigned to either the Stress or the Relax condition, based on their gender. This method of assignment resulted in four groups being formed: Stress Females (SF; $n = 33$), Stress Males (SM; $n = 21$), Relax Females (RF; $n = 25$) and Relax Males (RM; $n = 21$).

Materials

Participant self-report measurements.

Beck Depression Inventory-II (BDI-II). The BDI-II (Beck, Steer, & Brown, 1996) is a self-report questionnaire consisting of 21 items, each with a heading in order to alert the participant to its focus (Dozois, Dobson, & Ahnberg, 1998). Every item has four possible responses, each indicating a different degree of possible depressive symptomatology. Respondents are asked to choose the response that best suits how they have felt for the previous 2 weeks, with higher scores indicating greater levels of depression, with the distinct ranges being: 0 – 13, minimally depressed; 14 – 19, mildly depressed; 20 – 28, moderately depressed; 29 – 63, severely depressed (Beck et al., 1996).

The BDI-II was developed in order to comply with the DSM-IV's (APA, 1994) criteria for depression (Dozois et al., 1998). It also has good psychometric properties in that it is highly internally consistent (α = .91; Dozois et al., 1998) and has good test-retest reliability (α = .93; Beck et al., 1996).

For the purposes of this study, the BDI-II was used as a screening measure, with participants scoring 29 or above being excluded from the study[2].

State-Trait Anxiety Inventory (STAI). The STAI (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) consists of two parts. Form Y-1 measures an individual's anxiety at a specific point in time (state anxiety), while Form Y-2 is an indicator of general levels of anxiety (trait anxiety). Each form consists of 20 statements. These are measured on a Likert-type scale with a possible answers ranging from "not at all" (Form Y-1) or "almost never" (Form Y-2) to "very much so" (Form Y-1) or "almost always" (Form Y-2). Some of the items are reverse scored which helps to reduce response sets. The STAI has good psychometric properties in that it has a reliable factor structure, is highly internally consistent, and has high levels of validity (Spielberger & Vagg, 1984).

For the purposes of this study, the STAI – Trait was used as a control measure to establish the participants' general levels of anxiety in order to ensure that the groups experienced similar levels of anxiety in their everyday lives. The STAI – State was used to measure changes

[2] All participants who were excluded based on this criterion were given the contact details for student wellness so that they could receive counselling if they wished.

in participant self-report anxiety throughout Day 2 of the study. These scores were compared to the physiological data that was collected.

Physiological measurements.

Salivary cortisol collection and measurement. Cortisol was collected by means of saliva samples. These samples are an easy, effective, and non-intrusive way to collect cortisol and do not cause any stress for the participant (Garde & Hansen, 2005).

SARSTEDT Salivette® Cortisol swabs (Sarstedt, Nümbrecht, Germany) were used to collect the samples. Participants were required to chew the provided swabs for one minute. Once the samples were collected, they were immediately stored in individual, labelled tubes and were frozen until they could be transported to the National Health Services Laboratory at Groote Schuur Hospital, where they were analysed.

The saliva samples were taken at three points during the session: (a) at the beginning of the session, (b) 5 minutes after the participants had completed the relevant experimental manipulation (i.e., 25 minutes from onset of the stressor), and (c) 5 minutes after end of the memory testing. Maximum cortisol increases in response to a psychosocial stressor are generally seen at between 10 and 30 minutes after the end of the stressor, with a return to baseline levels usually occurring between 60 and 90 minutes after the end of the stressor (Kudielka & Kirschbaum, 2005). Thus, in this study salivary cortisol peaks should have been experienced during the cognitive testing phase, and levels should have returned to normal toward the end of the study.

Heart rate measurements. ECG measurements were taken throughout Day 2's session using a Vrije Universiteit Ambulatory Monitoring System version 5fs (VU-AMS; Vrije Universiteit, Amsterdam, Holland) device. This non-invasive device is portable, and participants were thus able to move around and walk between the two venues used in the study while wearing it.

The device was fitted at the beginning of Day 2's session. Following fitting, 5- minutes were allowed in order for the participants' heart rates to stabilise. Average heart rate measurements were taken from each of the following periods: (a) a 2-minute baseline measurement immediately following the 5-minute stabilization period referred to above, (b) the

final 10 minutes of the experimental manipulation, and (c) the 5 minutes following the end of the memory testing.

The acute psychosocial stressor.

The method used to induce stress in this study was a variation on the Trier Social Stress Test (TSST; Kirschbaum, Pirke, & Hellhammer, 1993), a psychosocial stressor involving a public speaking and mental arithmetic task. According to its developers, the TSST successfully and reliably induces increases in cortisol levels (Kirschbaum et al., 1993). Independent verification of these claims emerged from a meta-analytic study investigating under what conditions psychological stress results in increased cortisol levels (Dickerson & Kemeny, 2004). This meta-analysis study found that stressors which combined the two elements found in the TSST showed greater increases in cortisol compared to other psychosocial stressors. Dickerson and Kemeny (2004) explain this finding by the fact that TSST-type tasks include a social judgement aspect (where participants experience the possibility of having their performance being negatively evaluated), as well as the fact that these tasks are unpredictable (therefore participants do not feel completely in control of the situation). In addition, the TSST's ability to elicit quite large cortisol responses has been demonstrated in recent individual empirical studies (e.g., Luethi et al., 2009; Nater et al., 2007; Schoofs et al., 2008).

Participants were told that they needed to pretend that they were going for an interview at a job of their choice and that they would need to prepare a 5-minute speech detailing their suitability for that job. They were told that they would present this speech to an interviewing panel that would, with the help of a video-recording of the presentation, be analysing the content of their speech and their verbal and non-verbal behaviour.

The participants were then given 10 minutes to prepare their speech, using pen and paper. When the 10 minutes were over, they were taken to a room illuminated by a harsh, bright light which contained a video-camera and the interviewing panel (one male and one female), who were seated behind a desk. Participants were told that they were to stand in front of the panel in range of the video camera, and that they would need to speak for a full 5 minutes without the use of their notes, which were taken from them at this point. The researcher then stood behind the participant and timed their speech. If the participants finished their speeches before the allotted amount of time had elapsed, then they were told that time still remained and that they should

continue. If they stopped again before the 5-minute time limit, was finished, then they were asked a series of preset questions, namely: (a) what is the most difficult experience that you have had which will help you in this job? (b) what are some of your weaknesses? and (c) for what reasons should we not take you for this job? Once the full 5 minutes had elapsed, the researcher indicated this to the panel by waving her hand. The panel would then interrupt the participant and instruct him/her to perform a serial subtraction task (to minus the number 13 from 1022 repeatedly). Every time the participant made a mistake, he/she was told to stop and start again from 1022. Participants performed this task for 5 minutes. All questions and instructions were asked or given by the individual on the panel who was of the same sex as the participant undergoing the manipulation.

Memory tasks.

Declarative memory tasks.

Cued recall. Following de Quervain et al. (2003) and Lupien et al. (1999), I used a VPA cued recall task to test DM. My task was based on Uttl, Graf, and Richter's (2002) VPA-15, which in turn is similar to the one presented in the Wechsler Memory Scale – Third Revision (WMS-III; Psychological Corporation, 1998). The latter VPA task is described by Uttl et al. (*p*. 567) as being "among the most widely used instruments for assessing explicit episodic memory". The VPA-15 list contains 15 word pairs, of which four are regarded as 'related/easy' pairs (e.g., fruit-apple) and 11 are regarded as 'unrelated/difficult' pairs (e.g., bank-milk). The four related/easy word pairs and four of the unrelated/difficult pairs are taken directly from the Wechsler Memory Scale – Revised (Wechsler, 1987). For the purposes of the current study, 3 extra 'easy' pairs were added to Uttl et al.'s (2002) list in order to bring more balance to the word list and to allow for better comparison between the number of difficult and easy pairs remembered. One of the extra pairs was taken from Lupien et al. (1999; prison-thief) and one was taken from Shimamura, Jurica, Mangels, Gershberg, and Knight (1995; lion-circus). The third was created for the purposes of this study by the experimenter (colour-blue).

On Day 1 of the experimental protocol, the list was presented to each participant twice, with each presentation being followed by a cued recall test. The presentation order used was based on the relevant "study order" (1 or 2) provided by Uttl et al. (2002, p. 573), with three extra easy pairs included at randomly-determined places. The recall order used was based on the

relevant "recall order" specified by Uttl et al. (2002, p. 573), again with the three extra word pairs included at randomly-determined places.

Participants were told that they were going to be read a series of word pairs, and that they should remember the pairs because they would be tested on them afterwards. The pairs were read one at a time, with a 3-second break between each pair. Participants were then told that the experimenter would read the first word of each pair out loud, and that they had to give the second word of that pair. Following the first presentation and recall of the list, participants were told that they were going to be read the list again, and were then going to be tested on it again afterwards. They were then given the second study order and the second recall order in the same manner as the first.

On Day 2 of the experimental protocol, there was no study phase for the word pairs. The participants were merely read the first word of each pair (in a standard order devised by the experimenter) and, as before, were required to give the second word for each pair.

The word pairs and order of presentation for both the study and testing phase on Day 1 and for the cued-recall testing phase on Day 2 can be found in Appendix A.

Recognition. Following the completion of Day 2's cued recall task, participants were required to engage in a recognition task based on the cued recall word pairs. Ninety word pairs were presented to the participant via a computer using E-prime software version 1.1 (Psychology Software Tools, 2002, Pittsburgh, Pennsylvania). These 90 pairs included all 18 of the previously studied pairs and 72 pairs constructed specifically for this study by finding an 'easy' and a 'difficult' partner for each of the first words in the original pairs (e.g., fruit-vegetable and fruit-holiday) as well as for each of the second words in the original pairs (e.g., pear-apple and ship-apple). The computer programme presented the word pairs in a random order for each participant. Participants were required to indicate, using the '1' or '2' keys on the computer keyboard, whether the presented pair was one of the previously studied 18 or not. There was no time-limit for a response from the participants.

The word pairs used in the Day 2 recognition task can be found in Appendix B.

Working memory task. This study used an *n*-back test involving three difficulty levels: a 0-back condition, a 1-back condition and a 3-back condition. The task closely followed that used by Schoofs et al. (2008), and addressed the need identified by them for WM performance under

conditions of stress to be investigated at an easier difficulty level than those used in their study (a 2- and 3-back condition).

This test was presented via a computer using E-prime programming software version 1.1 (Psychology Software Tools, 2002, Pittsburgh, Pennsylvania) and is modified from an *n*-back script downloaded from http://step.psy.cmu.edu/scripts-plus/.

In this task, a random series of letters were presented on the computer screen. Participants were required to hit the 'F' key on the keyboard if the presented letter was considered a 'target' and the 'J' key if the letter was considered a 'non-target'. In the 0-back condition, the letter 'X' was the target. In the 1- and 3-back conditions, a letter was a target if it was the same letter that was presented *n*-letters previously (i.e., 1 letter previously for the 1-back and 3 letters previously for the 3-back condition; see Figure 1 for a graphic depiction of these instructions).

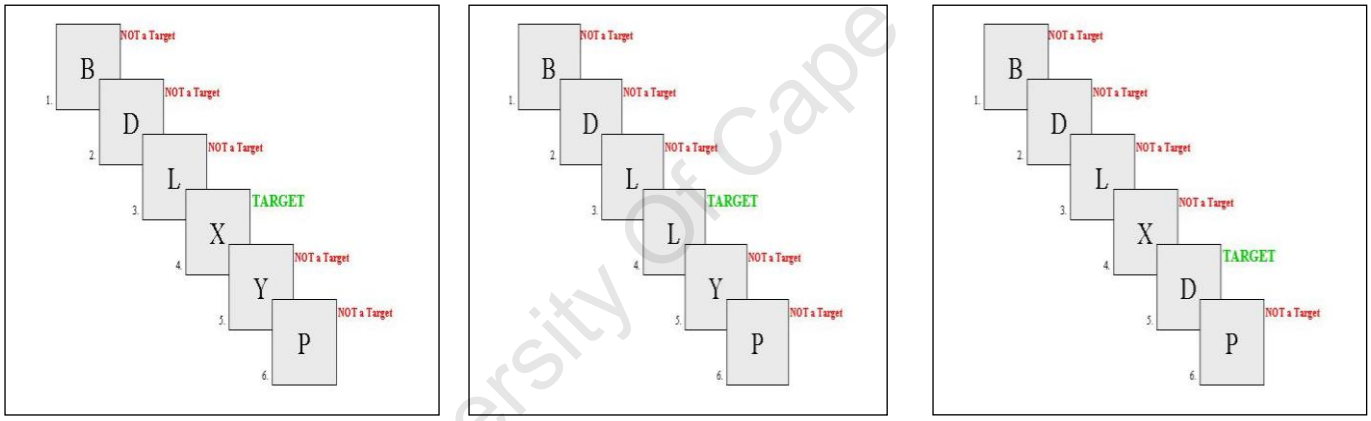

Figure 1. An example of the instructions for, from left to right, the 0-back, 1-back and 3-back conditions of the *n*-back WM task.

On Day 1, the participants completed a shortened version of the main task to be completed on Day 2 in order to ensure that they understood the instructions for each difficulty level. They were required to achieve an accuracy score of at least 70% on each difficulty level before they were allowed to move onto the next level. Thus, the participants repeated each difficulty level until they reached this goal. Each block of each difficulty level included 20 stimulus presentations and information about the percentage achieved was presented at the end of each level block.

On Day 2, the participants were required to complete one 0-back block followed by eight blocks alternating between the 1- and 3-back conditions. In each of these blocks there were 24

stimulus presentations, 33% of which were target stimuli. However, the first three letters of each block were non-targets and were not considered during data analysis. The letter 'X' was not used in any of the 1- or 3-back blocks, so as to prevent any confusion with the 0-back block. All of the information pertaining to the 1- and 3-back conditions follows the protocol of Schoofs et al. (2008), except for the time between the onset of one letter and the next (3518 ms), which is a little longer than the 2750 ms used by Schoofs et al. (2008).

Procedure

Day 1. As noted above, participants were pseudorandomly divided by gender into either a Stress group or a Relax group before meeting the researchers. On Day 1 of the experimental protocol, however, the procedure for all participants was identical. Each participant was met at the computer laboratory where the memory testing took place. Here they read and signed the consent form (Appendix C) and completed the BDI-II and the STAI – Trait questionnaires. The BDI-II was scored while the STAI – Trait was completed so that participants did not need to complete the rest of the study's protocol if they met the depression exclusion criterion. Once they had finished filling out the questionnaires, the two study and immediate cued-recall phases of the VPA test, as well as the 'practice' *n*-back test were administered. The VPA and *n*-back were administered in a counterbalanced order across participants to control for possible sequence effects.

At the end of this session, participants were reminded about their appointment for the next day, and were asked not to smoke, consume any food or drink, chew gum or engage in physical exercise for 2 hours before the start of their second session. This reminder was in line with protocols followed by other studies (e.g., Kirschbaum et al., 1993; Schoofs et al., 2008).

Day 2. On Day 2 of the experimental protocol, participants in all groups were met at the same venue that they had been tested in the previous day. Before testing commenced, they were reminded that they were free to terminate their participation in the study at any stage. The first STAI – State questionnaire was completed and the first saliva sample was taken. After this, the participants were fitted with the VU-AMS device and a 5-minute rest period was allowed for the device to normalise to the participants' heart rate, following which a 2-minute baseline reading was taken. From here the procedure differed depending on the group to which the participant had

been assigned. Those in the Stress condition underwent the 20-minute TSST procedure while those in the Relax condition sat in a comfortable chair, read non evocative magazines, and listened to relaxing music for 20 minutes.

Once the participants had completed their respective activities, they underwent a 5-minute relaxation period. After this, the experimenter collected a second saliva sample from the participants and had them complete a second STAI – State questionnaire. The second memory testing session then began.

The memory tests were completed in the same order as they had been administered in the first session. After testing was complete, participants had another relaxation period of 5-minutes, after which the VU-AMS was removed, a third saliva sample was taken and a third STAI-State was completed. The participants were then debriefed as to the purpose of the study; in particular, those in the Stress condition had the TSST explained to them so that they understood that they were not really being evaluated and that their performance was not really being recorded. All participants were asked not to discuss any aspect of the study with anyone else so as not to confound the results. Additionally, the female participants were asked to contact the experimenter of the first day of their next period.

Statistical Analysis

All statistical analyses were completed using STATISTICA version 8.0 (StatSoft, Inc., 2008, Tulsa, Oklahoma). The study's design allowed for both within- and between-groups analyses. The level for statistical significance was set at $\alpha = 0.05$.

Details about specific analyses are provided before presentation of their results. Unless otherwise stated, all of the required assumptions were upheld for each statistical analysis.

Results

Final Sample

Although a large number of participants were enrolled in the study, there was, unfortunately, a high rate of attrition during the data collection process.

Reasons for participant exclusion. Figure 2 is a diagrammatic explanation of the sources of attrition during the data collection process.

Participants excluded before the end of Day 2's session. Fifteen of the participants enrolled in the study did not complete it. Seven participants (three participants from the SF group, two participants from the SM group, and one participant from each of the RF and RM groups) were excluded from the study at the start of the first session on the basis of the BDI-II exclusion criterion. Two participants from the SF group and one participant from each of the SM, RF and RM groups did not arrive for their second session. One female in each group began their period before the start of their second session, and so did not complete the study. Finally, one participant from the SF group asked to discontinue her participation in the study during the experimental manipulation.

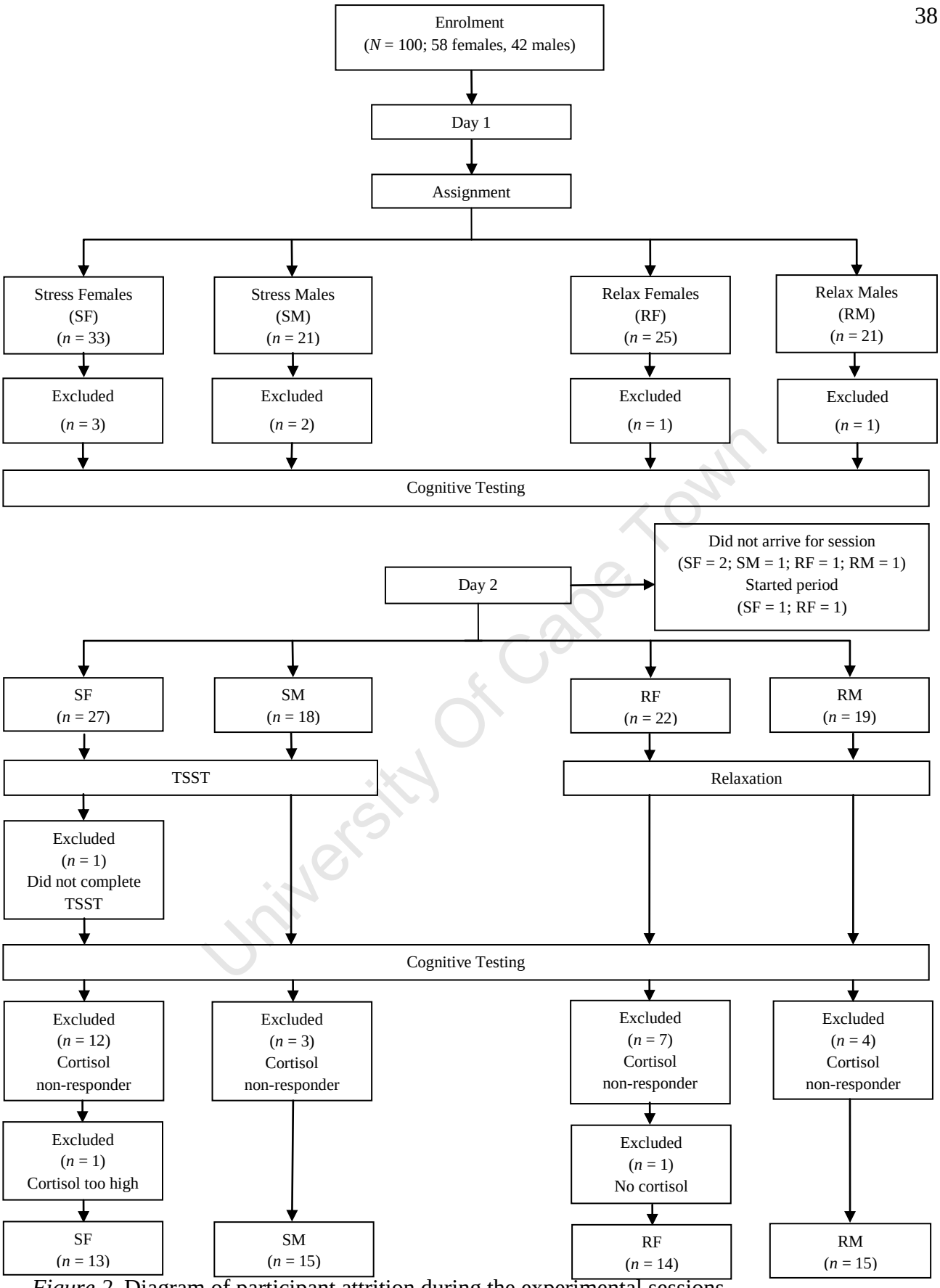

Figure 2. Diagram of participant attrition during the experimental sessions.

Cortisol responders vs. cortisol non-responders. The biggest reason for the reduced final sample is due to participants' cortisol levels not behaving as predicted. This is not an unusual finding as recent studies in this field have demonstrated that participants can generally be divided into 'cortisol responders' (i.e., those whose cortisol levels are raised in response to a stressor) and 'cortisol non-responders' (i.e., those whose cortisol levels remain the same or decrease in response to a stressor). Because the general premise of studies investigating the effects of stress on memory is that it is the increased cortisol levels that result from stress that affect memory functioning, there is a precedent in the literature to analyse the data from the cortisol responder's separately from that of the cortisol non-responders (see, e.g., Buchanan & Tranel, 2008; Elzinga & Roelofs, 2005). Studies utilizing this strategy have found that cortisol non-responders show a similar (Elzinga & Roelofs, 2005) or in some cases slightly better (Buchanan & Tranel, 2008) performance on the relevant memory tasks than non-stressed controls, who always show better performance than the cortisol responders.

Based on these findings, and because this study specifically aimed to investigate the effects of the HPA axis response to stress (i.e., increased cortisol levels) on memory functioning, only the data from those participants in the Stress group who were cortisol responders were analysed. In addition, only the data from those participants in the Relax group whose cortisol levels remained the same or decreased after the experimental manipulation were included in the analysis. Thus, the data from 12 participants in the SF group, 3 participants in the SM group, 7 participants in the RF group and 4 participants in the RM groups were excluded from the final statistical analyses.

Furthermore, data for another participant from the RF group were excluded because, due to experimenter error, there was no cortisol data for her. Finally, data for one other participant from the SF group were excluded from the analyses because, although her cortisol levels increased as expected after the manipulation, these levels were much higher than any of the other participants in the study. Her baseline cortisol level (30.11 nmol/l) was more than 18 standard deviations away from the mean for the rest of the participants in the combined Stress groups (M = 1.62 ± 1.57 nmol/l). Her post-TSST level (71.74 nmol/l) was more than 15 standard deviations away from the mean for the rest of the participants in the combined Stress groups (M = 6.53 ± 4.11 nmol/l). Her end of study recorded level (10.15 nmol/l) was more than three standard deviations away from the mean for the rest of the participants in the combined Stress groups

(M = 2.40 ± 2.25 nmol/l). In terms of the inverted-U hypothesis of the relationship between stress induced increases in cortisol levels and cognitive performance (de Kloet et al., 1999), it is possible that such differences in cortisol levels between this one participant and the rest of the Stress group participants may have placed them on separate points of the inverted-U curve, thereby suggesting that her performance on the memory tasks may not be in line with the rest of the participants. Therefore, including this participant in the final data analysis may have confounded the study's results. See Appendix D for the differences in the changes in salivary cortisol from baseline to post-manipulation between the total and final sample groups.

Final sample characteristics. As a result of the attrition and exclusions outlined above, the final number of participants included in the data analysis was 58 (SF group: n = 13; SM group: n = 15; RF group: n = 14; RM group: n = 15).

In order to analyse the characteristics (which acted as control measures to ensure that all participants were drawn from a similar population) of the final sample, 2 x 2 (Experimental Condition x Gender) between-groups factorial ANOVAs were conducted on the data for the ages of the participants, their BDI-II scores and their STAI – Trait scores (see Table 3 for descriptive statistics). The BDI-II and STAI – Trait scores of the sample were also compared to the normative data for those tests by means of single-sample t-tests. In addition, the accuracy of the control for menstrual cycle phase was investigated using descriptive statistics.

Table 3
Descriptive Statistics for Final Sample Characteristics

Measure	Group			
	SF	SM	RF	RM
	n = 13	n = 15	n = 14	n = 15
Age	19.23 (1.17)	20.20 (2.57)	19.43 (1.83)	19.33 (1.88)
BDI-II	12.61 (6.58)	9.27(6.76)	14.64 (7.19)	11.33 (5.69)
STAI – Trait	38.38 (9.02)	36.53 (6.99)	46.00 (10.76)	41.07 (9.12)

Note. Data present are means with standard deviations in parentheses.

Age. The participants' ages ranged from 18 to 27 years (M = 19.56 ± 1.94). Although slight violations of the normality assumptions were indicated for the factorial ANOVA conducted on this data, these were not very severe and the fact that ANOVA is a robust statistical

test meant that this violation should not have had too great an impact on the results. The analysis did not show significant main effects for Experimental Condition, $F(1, 53) = 0.42$, $p = .521$, partial $\eta^2 = .01$, or Gender, $F(1, 53) = 0.71$, $p = .402$, partial $\eta^2 = .01$. It also did not show a significant interaction effect between these two variables, $F(1, 53) = 1.06$, $p = .308$, partial $\eta^2 = .02$. Thus, it is clear that the average ages of the participants in the four groups did not differ statistically significantly, which implies that the results of the study were not confounded by differences in age between the participants in different groups.

BDI-II scores. Based on the mean BDI-II scores for the four groups, it can be seen that the SF, SM and RM groups all fell in the 'minimal' depression range, while the RF group falls just into the 'mild' depression range, indicating low levels of depressive symptomatology for the groups (Beck et al., 1996). The factorial ANOVA did not show a significant main effect for Experimental Condition, $F(1, 53) = 1.38$, $p = .245$, partial $\eta^2 = .03$ or Gender, $F(1, 53) = 3.65$, $p = .061$, $\eta^2 = .07$, nor did it show an interaction effect between Gender and Experimental Condition, $F(1, 53) = 0.01$, $p = .991$, $\eta^2 < .01$. Thus, it is clear that the four groups did not differ statistically significantly in terms of depressive symptomatology, which implies that the results of the study were not confounded by differences in pre-existing emotional states between the groups.

In addition, the participants in this study ($M = 11.90 \pm 6.69$) did not differ statistically significantly, $t(56) = -0.75$, $p = .456$, $d = 0.08$, from the normative data supplied by the BDI-II manual for college students ($M = 12.56 \pm 9.93$; Beck et al., 1996), indicating that this sample is representative of the general population of tertiary education students.

STAI – Trait anxiety scores. Analysis of the sample's trait anxiety scores shows that there was a significant main effect of Experimental Condition, $F(1, 53) = 6.41$, $p = .014$, partial $\eta^2 = .11$, in the absence of both a main effect for Gender, $F(1, 53) = 2.00$, $p = .163$, partial $\eta^2 = .04$, and an interaction effect between Experimental Condition and Gender, $F(1, 53) = 0.41$, $p = .523$, partial $\eta^2 = .01$. More specifically, the Stress groups ($M = 37.39 \pm 7.90$) showed statistically significantly lower levels of trait anxiety than the Relax groups ($M = 43.45 \pm 10.08$). However, it appears from the interaction result that this difference is only noticeable between the two bigger experimental condition groups, and disappears when gender is also taken into account. Thus, it appears that the four groups are similar in terms of general anxiety levels.

In addition, as with the BDI-II scores, the STAI – Trait scores for the participants in this study suggested that this sample was representative of the general population of college students. Single sample t-tests showed that neither the female participants' scores (M = 42.33 ± 10.51) nor the male participants' scores (M = 38.80 ± 8.31) differed statistically significantly (females: $t(26) = 0.96$, $p = .348$, $d = 0.19$; males: $t(29) = 0.33$, $p = .744$, $d = 0.05$) from the normative data for these groups (females: M = 40.40 ± 10.15; males: M = 38.30 ± 9.18) supplied by the STAI manual (Spielberger et al., 1983).

Menstrual cycle phase. Post-experimental self-report verification showed that 23 of the 27 female participants in the final sample (SF: n = 11; RF: n = 12) were tested in the desired phase of their menstrual cycles. The remaining four (two in each group) all took part in Day 2 of the study more than 6 days before the start of their menstrual cycle. Figure 3 shows the distribution of final sample female participants in the correct and incorrect phases of their menstrual cycles on Day 2 of the study by showing how many days away from the start of their period they were at the time of testing. Appendix E discusses the accuracy of menstrual cycle phase for the total female participant sample.

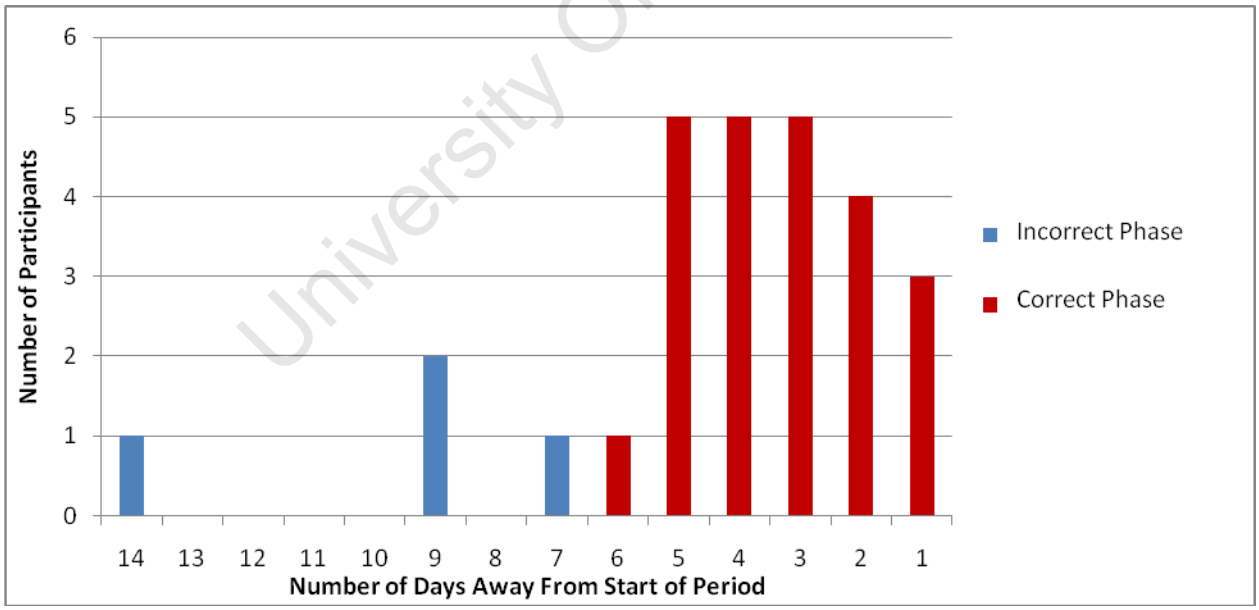

Figure 3. Number of final sample female participants in the correct and incorrect phases of their menstrual cycles on Day 2.

Experimental Manipulation

The analyses described in this section were conducted in order to test the effectiveness of the experimental manipulation (i.e., either the TSST or the relaxation period) on participants in the respective groups. For each of the relevant outcome variables, 2 x 2 x 3 (Experimental Condition [stress/relax] x Gender [male/female] x Stage of Testing Process [baseline/post-manipulation/end of session]) repeated measures ANOVAs were conducted and planned comparisons were run to test pre-existing hypotheses about where exactly between- and within-group differences would exist.

Table 4 provides descriptive statistics for each of the relevant self-report and physiological measurements (viz., STAI – State scores, salivary cortisol measurements, and heart rate data). It is important to note that due to hardware malfunctions, data were lost for some of the heart rate measurements. In some cases no data was recorded at all and in some cases only part of a specified time point was recorded or it was not clear which part of the data corresponded to which part of the session. In these cases, experimenter discretion was employed to decide which readings were included and which were not. For four heart rate measurements, data that were fully recorded were specifically removed due to the fact that the measurements appeared to be inaccurate ($M > 200$ bpm). Thus, the descriptive statistics presented are only for those participants who had usable data across all three measurement points.

Table 4
Descriptive Statistics for Self-Report and Physiological Measures

Measure	Group			
	SF	SM	RF	RM
	n = 13	*n* = 15	*n* = 14	*n* = 15
STAI – State				
Baseline	35.69 (7.09)	34.40 (7.76)	38.21 (8.61)	34.80 (10.48)
Post-manipulation	48.62 (10.85)	42.33 (11.64)	31.79 (7.53)	28.80 (6.20)
End of Session	33.23 (7.18)	29.60 (7.70)	31.79 (8.50)	29.67 (6.06)
Cortisol Level				
Baseline	1.34 (1.23)	1.86 (1.82)	1.21 (1.27)	1.95 (2.60)
Post-manipulation	4.99 (2.82)	7.86 (4.66)	0.99 (1.23)	1.27 (1.86)
End of Session	1.37 (0.87)	3.30 (2.69)	0.80 (0.91)	0.83 (0.78)
Heart Rate				
Baseline	81.26 (4.71)[a]	71.24 (12.58)[b]	77.61 (10.60)[a]	77.23 (17.16)[b]
Post-manipulation	118.51 (13.19)[a]	95.70 (15.94)[b]	73.99 (9.26)[a]	69.52 (13.07)[b]
End of Session	78.69 (7.32)[a]	71.21 (9.61)[b]	75.76 (11.15)[a]	69.18 (11.39)[b]

Note. Mean scores are provided with standard deviations in parentheses. Cortisol levels are measured in nanomoles per litre (nmol/l). Where cortisol levels for a participant were indicated to be < 0.50 nmol/l, 0.45 nmol/l was used as an estimate. Heart rate levels are measured in beats per minute (bpm).
[a]*n* = 7. [b]*n* = 12.

Self-report anxiety measure: STAI – State. The analysis showed statistically significant main effects for Experimental Condition, $F(1, 53) = 6.30$, $p = .015$, partial $\eta^2 = .12$ (Stress condition: $M = 37.31 \pm 1.37$; Relax condition: $M = 32.51 \pm 1.34$), and Stage of Testing Process, $F(2, 106) = 23.16$, $p < .001$, partial $\eta^2 = .30$ (Baseline: $M = 35.78 \pm 1.15$; Post-manipulation: $M = 37.88 \pm 1.23$; End of session: $M = 31.07 \pm 0.98$), with an absence of a main effect for Gender, $F(1, 53) = 2.95$, $p = .092$, partial $\eta^2 = .05$. In addition, there was, as expected, a significant interaction effect between the Experimental Condition and the Stage of the Testing Process, $F(2, 106) = 39.00$, $p < .001$, $\eta^2 = .42$, in the absence of statistically significant Experimental Condition x Gender, $F(1, 53) = 0.06$, $p = .816$, partial $\eta^2 < .01$, Gender x Stage of Testing Process, $F(2, 106) = 0.68$, $p = .509$, partial $\eta^2 = .01$, and Experimental Condition x Gender x Stage of Testing Process, $F(2, 106) = 0.91$, $p = .407$, partial $\eta^2 = .02$, interaction effects.

These results suggest that the experimental manipulation clearly had an effect on the subjective anxiety levels of the participants and that the gender variable was not a contributing factor to the changes in those levels. Thus, the Experimental Condition x Stage of Testing Process interaction effect was more closely examined using planned contrasts. Figure 4 shows the fluctuations in the participants' anxiety levels across Day 2.

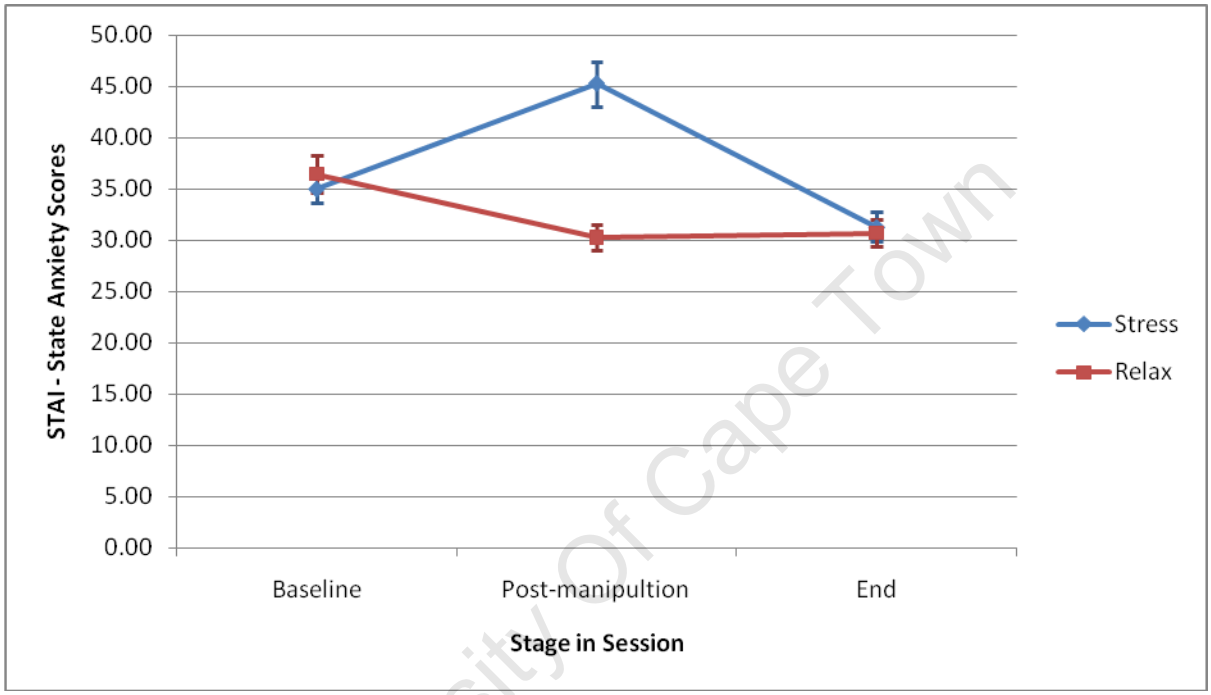

Figure 4. Changes in self-reported state anxiety levels on Day 2 for the combined Stress and combined Realx groups.Error bars indicate standard error of means.

The planned contrasts showed that the baseline anxiety levels of the Stress groups and of the Relax groups were not statistically significantly different, $F(1, 53) = 0.41$, $p = .526$. This result suggests that the groups entered Day 2 of testing with similar levels of state anxiety and that any changes in these levels after the experimental manipulation would therefore be comparable across the groups. In addition, single-sample t-tests showed that, at baseline, neither the female participants ($M = 37.00 \pm 7.88$) nor the male participants ($M = 34.60 \pm 9.06$) differed statistically significantly (females: $t(26) = -1.16$, $p = .256$, $d = 0.15$; males: $t(29) = -1.13$, $p = .268$, $d = 0.19$) from the normative data supplied by the STAI manual (Spielberger et al., 1983) for state anxiety in college students (females: $M = 38.76 \pm 11.95$; males: $M = 36.47 \pm 10.02$).

Thus, it appears that this study's sample was representative of the general population of college students, and that they were not excessively anxious at the start of the Day 2 session.

Further planned comparisons indicated that participants in the Stress groups showed a statistically significant increase in anxiety levels, from baseline (M = 35.00 ± 7.35) to post-manipulation (M = 45.25 ± 11.52), $F(1, 53) = 45.19$, $p < .001$. In contrast, the participants in the Relax groups showed a statistically significant decrease in anxiety levels from baseline (M = 36.45 ± 9.61) to post-manipulation (M = 30.24 ± 6.92), $F(1, 53) = 16.68$, $p < .001$. These results show that the TSST procedure successfully increased subjective levels of anxiety in the participants who were exposed to it, while the relaxation condition successfully decreased subjective levels of anxiety in the participants who were exposed to it. Therefore, after the experimental manipulation, participants in the Stress and Relax groups differed statistically significantly from each other in terms of subjective levels of anxiety with those participants in the Stress groups being far more subjectively anxious than the participants in the Relax groups.

In addition, participants in the SF and SM groups did not differ statistically significantly after the experimental manipulation, $F(1, 53) = 3.19$, $p = .080$, indicating both groups experienced similar (statistically significant) increases in levels of subjective anxiety. Similarly, participants in the RF and RM did not differ statistically significantly after the experimental manipulation, $F(1, 53) = 0.747$, $p = .391$, indicating that these groups experienced similar (statistically significant) decreases in levels of subjective anxiety.

Ethically, it was important that the TSST procedure did not have lasting effects on the participants in the Stress groups; in other words I had to ensure that they did not leave the study still feeling anxious from their exposure to the experimental manipulation, but rather left in an affective state similar to that in which they arrived. An examination of the baseline and end-of-session mean scores for the participants in the SF and SM groups, makes it clear that the participants' end-of-session STAI – State anxiety scores were actually lower than their baseline scores. Thus, it is clear that the participants who were exposed to the TSST were not still in a subjectively stressed state at the end of the Day 2 session.

Physiological stress measures.

Salivary cortisol levels. Due to violations of the assumptions of normality, homogeneity of variances (indicated by Levene's test) and sphericity (indicated by Mauchly's test) required for a repeated-measures ANOVA, it was necessary for us to perform transformations on the data. Log transformations corrected for the violation of homogeneity of variances and served to make the data somewhat more normally distributed. In order to correct the sphericity violation, $\chi^2(2) = 8.38$, $p = .015$, I performed a Greenhouse-Geisser degrees of freedom correction ($\varepsilon = .87$) on the log transformed data.

The results showed statistically significant main effects for Experimental Condition, $F(1, 53) = 27.10$, $p < .001$, partial $\eta^2 = .33$, and Stage in the Testing Process, $F(1.74, 92.22) = 32.22$, $p < .001$, partial $\eta^2 = .38$, in the absence of a statistically significant main effect for Gender, $F(1, 53) = 2.07$, $p = .156$, partial $\eta^2 = .04$. In addition, there was a statistically significant interaction effect between Experimental Condition and Stage in the Testing Process, $F(1.74, 92.22) = 52.72$, $p < .001$, $\eta^2 = .50$. However, there were no significant interaction effects between Experimental Condition and Gender, $F(1, 53) = 0.87$, $p = .356$, partial $\eta^2 = .02$, or between Gender and the Stage in the Testing Process, $F(1.74, 92.22) = 0.66$, $p = .521$, partial $\eta^2 = .01$, nor was there an Experimental Condition x Gender x Stage in the Testing Process interaction effect, $F(1.74, 92.22) = 1.33$, $p = .270$, partial $\eta^2 = .02$.

Figure 5 shows the fluctuations in the participants' cortisol levels across Day 2.

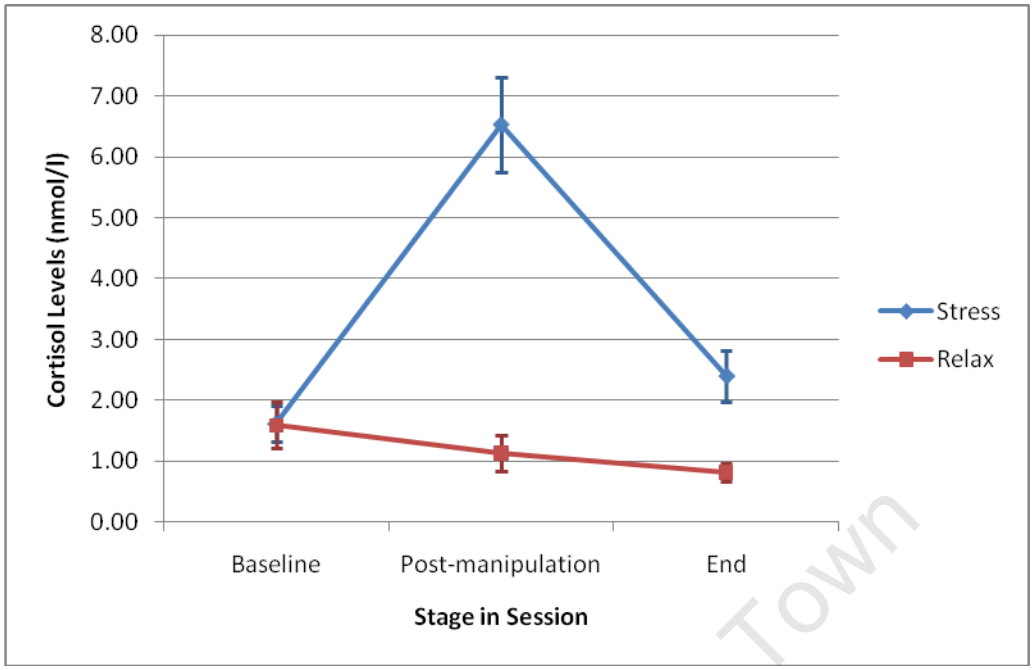

Figure 5. Changes in cortisol levels on Day 2 for the combined Stress and combined Relax groups. Error bars indicate standard error of means.

Similarly to the state anxiety scores, gender did not appear to have an effect on the changes in cortisol levels. Thus, planned contrasts were again only performed on the Experimental Condition and Stage in the Testing Process factors. One set of planned contrasts showed that participants in the Stress and Relax groups did not differ statistically significantly at baseline, $F(1, 53) = 0.23$, $p = .631$. This result suggests that any changes in participants' cortisol levels after the experimental manipulation would be comparable across the groups.

Further planned contrasts showed that participants in the Stress groups showed a statistically significant increase in their cortisol levels from baseline ($M = 1.62 \pm 1.57$) to post-manipulation ($M = 6.53 \pm 4.11$), $F(1, 53) = 141.03$, $p < .001$. At the same time, participants in the Relax groups showed a statistically significant decrease in their cortisol levels from baseline ($M = 1.59 \pm 2.07$) to post-manipulation ($M = 1.13 \pm 1.57$), $F(1, 53) = 5.10$, $p = .028$. Therefore, it is clear that (a) the TSST procedure successfully increased cortisol levels in the Stress groups, and (b) that post-manipulation, participants in the Stress and Relax groups differed statistically significantly from each other, with cortisol levels in the former being much higher than those in the latter.

It is important to note that participants in the SF and SM groups did differ statistically significantly after the experimental manipulation, $F(1, 53) = 6.50$, $p = .014$, with the participants

in the SM group experiencing far greater salivary cortisol levels post-manipulation. This result is in line with previous research which reports that male participants tend to experience higher cortisol levels post-manipulation than do female participants (e.g., Elzinga & Roelofs, 2005; Wolf, Schommer, et al., 2001). Nonetheless, this sex difference may need to be taken into account when interpreting performance on the memory tests.

Once again, it was important, for ethical reasons, to check that the participants in the Stress groups did not leave the study in a more stressed state than when they arrived. Planned comparison analyses of the Stress groups' cortisol levels at the end of the session ($M = 2.40 \pm 2.25$) showed that these were statistically significantly higher than their baseline levels ($M = 1.62 \pm 1.57$), $F(1, 53) = 7.53$, $p = .008$. However, based on the means presented in Table 4, it is clear that the participants in the SF groups' cortisol levels were only a little higher at the end of the session than at baseline, and a planned contrast showed that this difference was not statistically significant, $F(1, 53) = 0.55$, $p = .464$. Thus, it was only the SM participants who showed these statistically significantly greater levels at the end of the session, $F(1, 53) = 10.47$, $p = .002$. Although this result does suggest that the participants in the SM group were more stressed at the end of the session than at the beginning, they did show a statistically significant decrease in cortisol from after the TSST to the end of the study, $F(1, 53) = 50.92$, $p < .001$. Kudielka and Kirschbaum (2005) state it usually takes between 60- and 90-minutes for baseline cortisol levels to be returned to after the end of the stressor. In the current study, the point at which the third saliva sample was collected was only about 60 minutes after the end of the TSST. Therefore, while the participants in the SF group, who had a smaller stress response, had already returned to almost baseline, the participants in the SM group, who had a larger stress response, may not have reached this point yet. It is most likely that given a little more time, their cortisol levels would have returned to baseline.

Heart rate measurements. Due to the data violating the assumptions of normality, homogeneity of variances (indicated by Levene's test) and of sphericity (indicated by Mauchley's test), log transformations were performed on the data. Although this transformation appeared to make the violation of homogeneity of variances worse, it did make the data a little more normally distributed, and also fixed the sphericity violation. Thus, the log transformed data were used for the analysis, but the results still need to be interpreted with some caution.

The analysis showed significant main effects for Experimental Condition, $F(1, 34) = 7.80$, $p = .009$, partial $\eta^2 = .19$, Gender, $F(1, 34) = 4.59$, $p = .039$, partial $\eta^2 = .12$ and Stage of the Testing Process, $F(2, 68) = 62.17$, $p < .001$, partial $\eta^2 = .65$. In addition, there was a statistically significant interaction effect for Experimental Condition and Stage of the Testing Process, $F(2, 68) = 97.17$, $p < .001$, partial $\eta^2 = .74$. However, the Experimental Condition x Gender, $F(1, 34) = 0.86$, $p = .360$, partial $\eta^2 = .03$, Gender x Stage in the Testing Process, $F(2, 68) = 2.19$, $p = .120$, partial $\eta^2 = .06$, and Experimental Condition x Gender x Stage of the Testing Process, $F(2, 68) = 2.74$, $p = .071$, partial $\eta^2 = .08$ interaction effects were not statistically significant. Figure 6 shows the fluctuations in the participants' heart rate levels across Day 2.

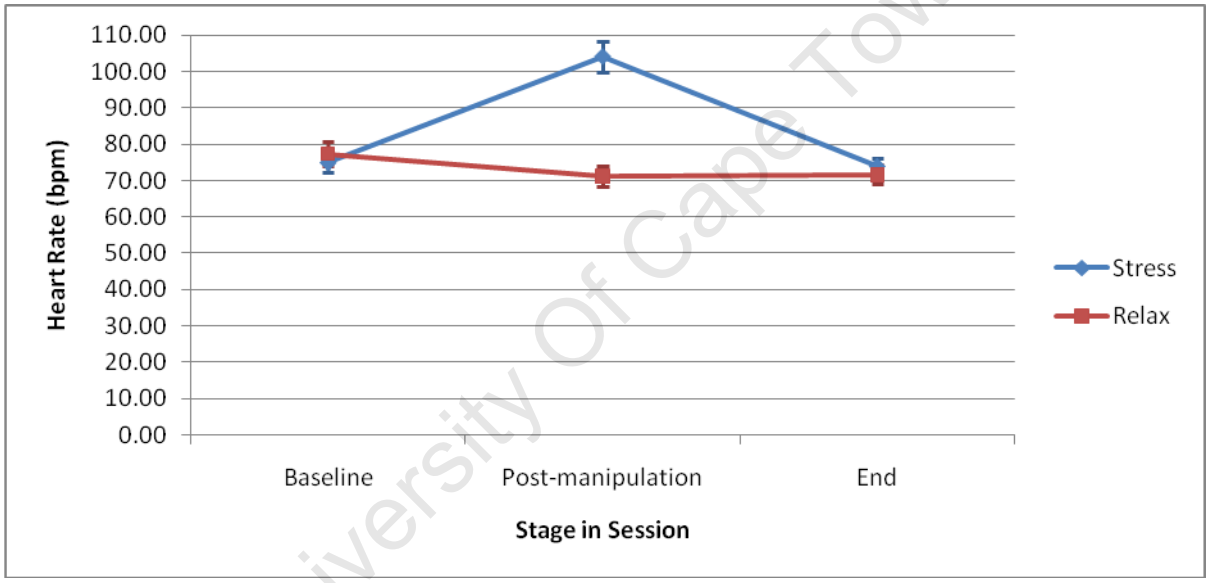

Figure 6. Changes in heart rate levels across Day 2 for the combined Stress and combined Relax groups. Error bars indicate standard error of means.

Although the main effect for Gender showed that the female participants ($M = 84.30 \pm 3.01$) had an overall faster heart rate than the male participants ($M = 75.68 \pm 2.30$), gender did not play a role in any of the interaction effects. Thus, the planned contrasts once again focused on the Experimental Condition x Stage in the Testing Process interaction. One set of planned contrasts showed that participants in the four groups did not differ statistically significantly in terms of their baseline heart rate levels, $F(1, 34) = 0.02$, $p = .879$. Further planned contrasts showed those participants in the Stress groups experienced a statistically significant increase in heart rate from baseline ($M = 74.93 \pm 11.35$) to post-TSST ($M = 104.11 \pm 18.47$), $F(1, 34) =$

179.73, $p < .001$. In contrast, those participants in the Relax groups showed a statistically significant decrease in heart rate from baseline ($M = 77.37 \pm 14.75$) to post-relaxation ($M = 71.16 \pm 11.74$), $F(1, 34) = 8.38$, $p = .007$. It is therefore clear that the TSST procedure successfully increased heart rates while the relaxation period successfully decreased them, leaving the Stress participants and the Relax participants with statistically significantly different heart rate levels after the experimental manipulation.

In addition, it is clear from Table 4 that participants in all four groups showed lower heart rate levels at the end of the session compared to at baseline. These values suggest that all participants were in a more relaxed state at the end of the session than at the beginning of it, once again suggesting that there were no long-term effects of the stressor on the participants.

In summary, it is clear from the converging data of the self-report anxiety measure, the salivary cortisol measure (representing HPA axis functioning), and the cardiovascular measure that the experimental manipulation worked as intended. Therefore, I successfully induced a stress response in the participants in the Stress groups, while reducing stress indicators in the participants in the Relax groups.

Memory Tasks

Declarative memory.

Cued recall task. The cued recall DM task was scored based on how many word pairs were correctly recalled. Slight variations from the original words were scored as correct (e.g., 'cry' for 'cries'). In line with Kuhlmann et al. (2005, p. 2978), "[t]o account for possible within- and between-subject variance in initial learning" the score on the cued recall task on Day 2 (i.e., the primary dependent variable under consideration here) "was expressed as the percentage of words remembered in relation to the second (and last) learning trial on [Day 1]". The outcome of this calculation is referred to here as the 'percentage savings' score for cued-recall DM.

In order to compare the between- and within- group results for the four groups for both days, 2 x 2 (Experimental Condition x Gender) factorial ANOVAs and 2 x 2 x 2 (Experimental Condition x Gender x Time) repeated-measures ANOVAs were conducted. Table 5 shows the descriptive statistics for performance over the three cued-recall trials, and for the percentage savings outcome variable. Figure 7 shows the groups' mean total scores across the three trials.

Table 5

Descriptive Statistics for Cued Recall DM Test Scores

Measure	Group			
	SF	SM	RF	RM
	n = 13	*n* = 15	*n* = 14	*n* = 15
VPA-IR-1				
Easy Pairs[a]	6.08 (1.04)	5.67 (1.23)	5.79 (1.53)	4.60 (1.35)
Difficult Pairs[b]	4.23 (2.56)	4.00 (2.73)	4.21 (3.38)	3.47 (2.85)
Total Pairs[c]	10.31 (3.25)	9.67 (3.58)	10.00 (4.46)	8.07 (3.62)
VPA-IR-2				
Easy Pairs[a]	6.77 (0.44)	6.67 (0.62)	6.71 (0.47)	6.40 (0.91)
Difficult Pairs[b]	7.85 (2.82)	7.73 (2.37)	6.93 (3.45)	5.40 (2.72)
Total Pairs[c]	14.62 (3.02)	14.40 (2.61)	13.64 (3.67)	11.80 (3.28)
VPA-24DR				
Easy Pairs[a]	6.31 (0.75)	6.00 (1.07)	6.07 (1.14)	5.53 (0.99)
Difficult Pairs[b]	5.08 (3.50)	4.93 (2.63)	4.57 (3.20)	3.67 (2.87)
Total Pairs[c]	11.39 (3.99)	10.93 (3.33)	10.64 (3.77)	9.20 (3.63)
Percentage Savings: VPA-24DR / VPA-IR-2				
Easy Pairs	93.22 (9.62)	89.62 (11.35)	90.14 (14.00)	87.54 (16.60)
Difficult Pairs	58.44 (24.25)	60.78 (18.14)	60.25 (27.94)	62.47 (38.92)
Total Pairs	76.09 (13.26)	74.93 (12.67)	76.88 (11.14)	77.82 (19.76)

Note. Data presented are means with standard deviations in parentheses. VPA-IR-1 refers to the first immediate cued recall trial on Day 1. VPA-IR-2 refers to the second immediate cued recall trial on Day 2. VPA-24DR refers to the delayed cued recall trial on Day 2.

[a]Maximum possible recall for Easy Pairs = 7. [b]Maximum possible recall for Difficult Pairs = 11. [c]Maximum possible recall for Total Pairs = 18.

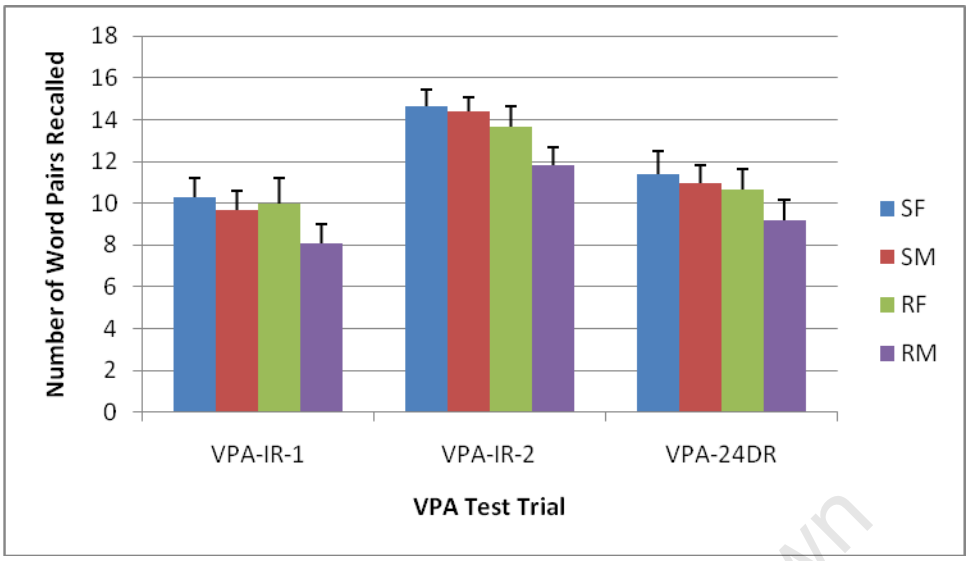

Figure 7. Average number of word pairs recalled by each group across the three VPA trials. Error bars indicate standard error of means.

Day 1: Total pairs. Results from the 2 x 2 factorial ANOVA conducted on the first immediate cued recall test (VPA-IR-1) data did not show statistically significant main effects for either Experimental Condition, $F(1, 53) = 0.92$, $p = .343$, partial $\eta^2 = .02$, or Gender, $F(1, 53) = 1.67$, $p = .202$, partial $\eta^2 = .03$, nor did they show a statistically significant Experimental Condition x Gender interaction effect, $F(1, 53) = 0.42$, $p = .519$, partial $\eta^2 = .01$. These results suggest that all participants (regardless of gender or group assignment) encoded and recalled a similar number of word pairs after the first presentation of the VPA list.

Results from the 2 x 2 factorial ANOVA conducted on the second immediate cued recall test (VPA-IR-2) data showed a statistically significant main effect for Experimental Condition, $F(1, 53) = 4.53$, $p = .038$, partial $\eta^2 = .08$, in the absence of a statistically significant main effect for Gender, $F(1, 53) = 1.50$, $p = .226$, partial $\eta^2 = .03$, or an Experimental Condition x Gender interaction effect, $F(1, 53) = 0.94$, $p = .337$, $\eta^2 = .02$. Although the Stress groups ($M = 14.50 \pm 2.76$) recalled more word pairs than the Relax groups ($M = 12.69 \pm 3.54$) after the second presentation of the word pair list, as indicated by the lack of statistical significance in the interaction effect, this difference no longer exists when all four groups are taken into account. In addition, the small effect size shows that even though there was a statistically significant result for the Experimental Condition main effect, the differences between the groups were not very large, again showing that this result should not have had too much of an effect on the Day 2 data.

The 2 x 2 x 2 repeated-measures factorial ANOVA conducted on the VPA-IR-1 and VPA-IR-2 data produced only one statistically significant result: the main effect of Time, $F(1, 53) = 216.86$, $p < .001$, partial $\eta^2 = .80$. This result indicates that a statistically significantly greater number of word pairs were recalled after VPA-IR-2 ($M = 13.62 \pm 0.42$) than after VPA-IR-1 ($M = 9.51 \pm 0.50$), and the effect size value shows that this difference is large, confirming that the second presentation of the word list benefitted the participants' encoding of the word pairs. However, the absence of significant main effects for Experimental Condition, $F(1, 53) = 2.44$, $p = .125$, partial $\eta^2 = .04$, and Gender, $F(1, 53) = 1.74$, $p = .193$, partial $\eta^2 = .03$, and the absence of significant interaction effects between Experimental Condition and Gender, $F(1, 53) = 0.69$, $p = .410$, partial $\eta^2 = .01$, Experimental Condition and Time, $F(1, 53) = 2.23$, $p = .141$, partial $\eta^2 = .04$, Gender and Time, $F(1, 53) = 0.21$, $p = .65$, partial $\eta^2 < .01$, and Experimental Condition x Gender x Time, $F(1, 53) = 0.09$, $p = .765$, partial $\eta^2 < .01$, show that this benefit to encoding did not differ across the four groups. These results also suggest that all participants across the four groups therefore showed similar increases in cued recall performance from VPA-IR-1 to VPA-IR-2.

Overall, the results presented thus far in this section suggest that all four groups entered Day 2's session having encoded a similar number of word pairs; hence, any between-group differences present after the experimental manipulation cannot be attributed to differential rates of encoding on Day 1. However, even the small differences in encoding between the four groups seen after VPA-IR-2 were further controlled for by using within-groups analyses of the Day 2 data, which included the calculation of the percentage savings score for each group. Therefore, Day 2's data were analysed using both between- and within- group analyses.

Day 2: Total pairs. The 2 x 2 factorial ANOVA conducted on the delayed cued recall test (VPA-2DR) results did not show any statistically significant results (Experimental Condition: $F(1, 53) = 1.61$, $p = .210$, partial $\eta^2 = .03$; Gender: $F(1, 53) = 0.94$, $p = .336$, partial $\eta^2 = .02$; Experimental Condition x Gender: $F(1, 53) = 0.26$, $p = .613$, partial $\eta^2 = .01$).

The 2 x 2 x 2 repeated-measures ANOVA conducted on the results from the VPA-24DR and VPA-IR-2 showed a statistically significant main effect of Time, $F(1, 53) = 167.73$, $p < .001$, partial $\eta^2 = .76$, in the absence of significant main effects of Experimental Condition, $F(1, 53) = 2.96$, $p = .091$, partial $\eta^2 = .05$, and Gender, $F(1, 53) = 1.27$, $p = .266$, partial $\eta^2 = .02$. In addition, there were no statistically significant interaction effects for this test (Experimental Condition x

Gender: $F(1, 53) = 0.56$, $p = .459$, partial $\eta^2 = .01$; Experimental Condition x Time: $F(1, 53) = 1.34$; $p = .253$, partial $\eta^2 = .03$; Gender x Time: $F(1, 53) = 0.03$, $p = .864$, partial $\eta^2 < .01$; Experimental Condition x Gender x Time: $F(1, 53) = 0.45$, $p = .506$, partial $\eta^2 = .01$).

Taken together, the results presented in this section show that there was a large decrease in the number of words recalled on Day 2 ($M = 10.54 \pm 0.49$) from the number recalled from the second recall test on Day 1($M = 13.62 \pm 0.42$), indicated by both the statistically significant main effect of Time in the 2 x 2 x 2 ANOVA and the corresponding large effect size. The lack of any other statistically significant results or large effect sizes, however, suggests that this decrease was experienced by all participants, regardless of experimental condition or gender. Additional independent sample t-tests were performed to double check that these differences were in fact seen in participants in all four of the groups. The results of these analyses confirmed that participants in all four of the groups did experience statistically significant decreases in cued recall performance on Day 2 (SF: $p = .029$; SM: $p = .004$; RF: $p = .043$; RM: $p = .049$). Therefore, it is difficult to attribute these decreased performances to anything other than the effects of time.

Percentage savings score. This analysis was run in order to investigate within-subject changes in cued recall from Day 1 to Day 2. Results for the Levene's test of homogeneity of variances showed that the data to be used in the 2 x 2 factorial ANOVA analysis violated this assumption ($p = .023$). Therefore, I performed log transformations on the data. Although this did not entirely correct this assumption violation, it did improve it ($p = .041$) and resulted in the data being slightly more normally distributed than before. These corrections and the fact that ANOVA is a robust statistical test mean that the results should not have been too affected by the Levene's violation.

The results of the subsequent ANOVA did not show significant main effects for either Experimental Condition, $F(1, 53) = 0.12$, $p = .731$, partial $\eta^2 < .01$, or Gender, $F(1, 53) = 0.05$, $p = .829$, partial $\eta^2 < .01$. There was also no significant Experimental Condition x Gender interaction effect, $F(1, 53) = 0.01$, $p = 940$, partial $\eta^2 < .01$. Thus, it is clear that no differences existed between the participants in terms of how many of the encoded Day 1 word pairs were recalled on Day 2.

Recognition task. Performance on the recognition task was primarily assessed using a d-prime (*d'*) score as the dependent variable. This *d'* score was calculated based on the 'hit' (word

pairs correctly identified as being on the original VPA list) and 'false alarm' (FA; word pairs incorrectly identified as being on the original VPA list) rates for each participant. The *d*' statistic is calculated as *d'* = z(FA) – z(H) with bigger *d'* values indicating greater discrimination between the original and distracter stimuli, and therefore better performance on the test. For perfect hit or FA rates (1 or 0 respectively), the formula 1 – 1/(2N) was used to calculate adjusted hit rates, and the formula 1/(2N) was used to calculate adjusted FA rates(see http://psy.ucsd.edu/~kang/sdt/sdt.htm for more details). A 2 (Experimental Condition) x 2 (Gender) factorial ANOVA was used to compare *d'* scores.

In addition, ratio scores capturing the proportion of between the hits on the recognition task and the VPA-IR-2 and VPA-24DR scores (Hits / VPA-IR-2 and Hits / VPA-24DR, respectively)were calculated and 2 x 2 repeated-measures factorial ANOVAs were run on these data in order to investigate the differential effects of stress on cued recall versus recognition memory.

Furthermore, average reaction times (RTs) were also calculated for the hits, FAs, total correct responses and total incorrect responses on the recognition task. Again, 2 x 2 factorial ANOVAs were conducted on these data.

Although there were slight violations of the assumption of normality for most of the analyses in this section, ANOVA is a robust statistical test, and so the violations should not have had too much of an impact on the results.

Table 6
Descriptive Statistics for Recognition DM Test Scores

Measure	SF *n* = 13	SM *n* = 15	RF *n* = 14	RM *n* = 15
Hits[a]	16.23 (1.42)	15.53 (1.55)	15.14 (2.41)	14.00 (1.81)
False Alarms[b]	3.46 (4.67)	4.07 (4.20)	5.14 (7.45)	4.67 (3.90)
d'[c]	3.27 (0.92)	2.90 (0.83)	2.91 (1.09)	2.45 (0.74)
Hits / VPA-IR-2	114.59 (20.77)	110.27 (16.55)	116.34 (29.54)	125.21 (28.04)
Hits / VPA-24DR	161.13 (47.34)	151.62 (35.94)	155.02 (46.89)	171.40 (61.41)
RT				
Correct Response	1536.44 (260.08)	1781.18 (553.36)	1571.25 (395.63)	1880.43 (451.70)
Incorrect Response	2696.07 (1711.53)[d]	2909.09 (1367.51)[e]	2323.48 (1195.47)[f]	2557.88 (905.01)[e]

Note. Mean scores are provided with standard deviations in parentheses. Reaction times measured in milliseconds (ms).
[a]Maximum possible hits is 18. [b]Maximum possible false alarms is 72. [c]Maximum possible *d'* score is 4.38. [d]n = 10. [e]n = 14. [f]n = 12.

d' scores. Analysis of the *d'* scores showed no statistically significant main effects for Experimental Condition, $F(1, 53) = 2.88$, $p = .095$, partial $\eta^2 = .05$, or Gender, $F(1, 53) = 2.94$, $p = .092$, partial $\eta^2 = .05$, nor was there a statistically significant interaction effect between these two variables, $F(1, 53) = 0.04$, $p = .849$, partial $\eta^2 < .01$. Thus, participants in all four groups were discriminating equally well between the original word pairs and the distracter word pairs.

Ratio scores. The analysis of the recognition hits to VPA-24DR ratio scores did not show any statistically significant main or interaction effects (Experimental Condition: $F(1, 53) = 0.28$, $p = .468$, partial $\eta^2 = .01$; Gender: $F(1, 53) = 0.07$, $p = .792$, partial $\eta^2 < .01$; Experimental Condition x Gender, $F(1, 53) = 1.00$, $p = .322$, partial $\eta^2 = .02$). This result indicates that, as predicted, the participants in all four groups recognised more words on the recognition test than they recalled on the VPAI-24DR trial, although the participants in the different groups did not differ in the percentage of words recalled (see Table 6).

The analysis for the hits to VPA-IR-2 ratio scores also showed no significant main or interaction effects (Experimental condition: $F(1, 53) = 1.67$, $p = .202$, partial $\eta^2 = .03$; Gender: $F(1, 53) = 0.12$, $p = .726$, partial $\eta^2 < .01$; Experimental Condition x Gender: $F(1, 53) = 1.04$, $p = .312$, partial $\eta^2 = .02$). This result again shows that the participants in all four groups accurately

remembered more word pairs on the recognition test than they did on the VPA-IR-2 cued recall test, but that neither experimental group nor gender had an impact on these ratio scores (see Table 6).

Reaction times. The results for the ANOVA conducted on the correct response RTs showed a statistically significant main effect for Gender, $F(1, 53) = 5.78$, $p = .020$, partial $\eta^2 = .098$, in the absence of a statistically significant main effect for Experimental Condition, $F(1, 53) = 0.34$, $p = .563$, partial $\eta^2 = .01$, or an Experimental Condition x Gender interaction effect, $F(1, 53) = 0.08$, $p = .781$, partial $\eta^2 < .01$. The data showed that the male participants ($M = 1830.81 \pm 498.87$) responded more slowly than the female participants ($M = 1558.43 \pm 325.83$) when making correct choices about the presented word pairs, regardless of the experimental group into which they were assigned. Therefore, it appears that in this case RT was mediated by gender rather than by psychological state. Interestingly, the larger standard deviations in these RTs among the combined male participants than among the combined female participants suggests that there was a lot more variation between the male participants in making correct responses, possibly indicating more uncertainty among these participants. In addition, it is clear that participants in both of the Stress groups performed a little faster than the participants in the gender-equivalent Relax groups.

The results for the ANOVA conducted on the incorrect response RTs showed no statistically significant main or interaction effects (Experimental condition: $F(1, 53) = 0.96$, $p = .333$, partial $\eta^2 = .02$; Gender: $F(1, 53) = 0.37$, $p = .548$, partial $\eta^2 = .01$; Experimental Condition x Gender, $F(1, 53) < 0.01$, $p = .977$, partial $\eta^2 < .01$). In spite of this, it is interesting to note the much larger standard deviations for these data than for the correct RT data (see Table 6). The longer average RTs and larger standard deviations for these data indicate that there was a lot more variation within the groups in their RTs when making incorrect responses. This result shows that there was potentially more hesitation among participants, regardless of experimental group or gender, when making these choices.

Working memory. Because the version of the *n*-back task performed on Day 1 was only a check to make sure that the participants understood what was required of them, statistical analyses were not run on the data. Descriptive statistics are presented as an indication of the participants' performance.

For the Day 2 data, the number of 'hits' (target letters correctly identified) and 'correct rejections' (non-target letters correctly identified) were summed and a percentage of correct responses for each overall task difficulty (i.e., total 1-back and total 3-back), as well as for each individual block of each task difficulty (i.e., each individual 1-back block and each individual 3-back block), were obtained for each participant. I also calculated the mean reaction times (RTs) for the correct responses and for the incorrect responses. Following the analytic strategy used by Schoofs et al. (2008), I initially conducted 2 x 2 x 2 x 4 (Experimental Condition x Gender x Task Difficulty [1-back/3-back] x Block [Time]) repeated measures ANOVAs on the percentage correct responses and mean RTs outcome variables. In addition, 2 x 2 (Experimental Condition x Gender) factorial ANOVAs were conducted on the overall percentage correct responses for each difficulty level, on the correct response RTs, and on the incorrect response RTs. As previously stated, RTs are considered to be an important component of true WM tasks and should therefore be analysed (Schoofs et al., 2008).

Day 1. Participants were required to score at least 70% on each difficulty level before they could proceed with the task. The percentage of participants for each group who achieved the required score for each difficulty level on the first attempt is shown in Figure 8.

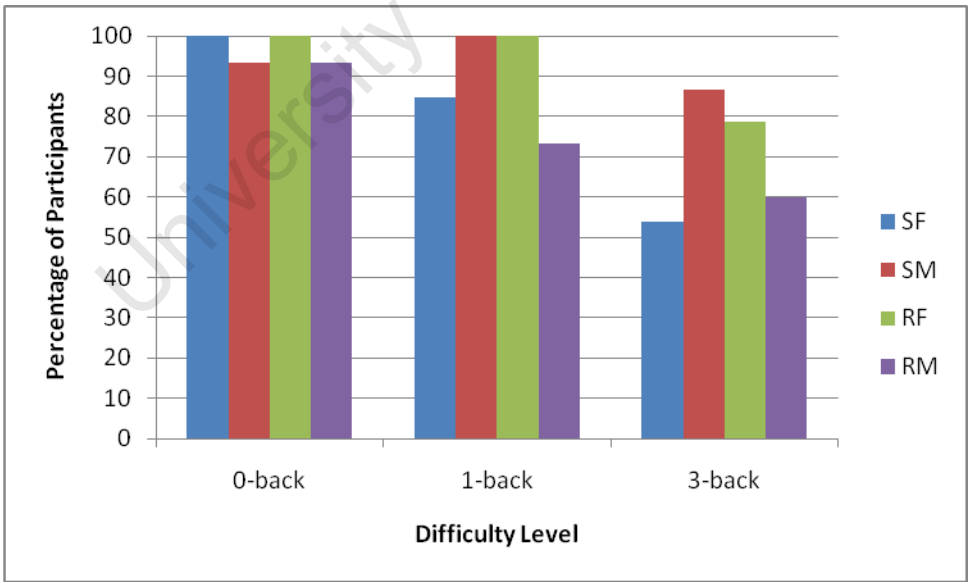

Figure 8. Percentge of participants attaining the minimum requirement (70%) on their first attempt.

0-back. Only two participants (SM: $n = 1$; RM: $n = 1$) did not meet the 70% criterion on the first attempt. However, it seemed from the data file that the participant in the SM group only failed this level because he confused the 'target' and 'non-target' keys. The participant in the RM group got the first 6 responses to the block incorrect and after that made just one more error. This pattern of data suggests that he may not have completely understood what was required of him during his first attempt at this level. Unsurprisingly then, both these participants met the required criterion on their second attempt. In addition, both of these participants met the required accuracy rate for both the 1- and 3-back difficulty levels on their first attempts. Therefore, overall it does not appear that the two participants who failed to meet the 70% criterion on their first attempt had any less aptitude for the task than the rest of the participants.

1-back. Of the 58 final participants, 51 achieved the required 70% accuracy on their first attempt. Only one participant, in the SF group, required three attempts to meet the criterion at this difficulty level. On closer analysis of her data, however, it was clear that there were a number of stimuli to which she was not responding on her first two attempts and that if she had responded correctly to these stimuli, she would have reached the 70% criterion on both attempts. On her third attempt, she only failed to respond to the first stimulus, which on Day 2 was discarded from the analysis.

3-back. Thirty-nine of the 57 participants achieved the required 70% accuracy on their first attempt, with an additional 13 participants achieving the criterion after two attempts. Only four participants did not achieve the criterion in two attempts.

Of those participants who did not achieve the criterion in two attempts, two participants in the RM group did not achieve the required 70% on this level due to experimenter error. One participant only completed this level once and achieved 50%, while the second completed this level four times but only managed to achieve a best accuracy score of 65%. As a result, it was necessary to check how these two participants' Day 2 performances compared with the others in their group in order to ensure that the fact that they did not reach the required proficiency did not constitute a possible confounding variable. The average overall 3-back percentage correct response for the other 13 RM participants was 84.80 ± 10.42, with the minimum being 61.90%. One of the participants achieved an overall score that was within one standard deviation of this mean (78.57%), indicating that his data could be retained, while the other participant scored just below one standard deviation away from this mean (72.62%), indicating that his data should

probably be excluded. However, when looking at these participants' percentage of correct responses for the first 3-back block, the first participant's score (71.43%) fell more than two standard deviations away from the mean score for the rest of the participants (M = 89.01 ± 7.87), while the second participants' score (61.90%) fell more than three standard deviations away from the mean score. Consequently, because there is no way of telling whether these scores are a true reflection of their ability, or as a result of their incomplete performance of the task on Day 1, these participants' data were removed from the final statistical analyses.

Day 2: Overall data analysis. In addition to the two participants in the RM group removed due to the factors outlined above, data were also removed for four other participants (SF: n = 1; RF: n = 1; RM: n = 2) due to a large number of no responses on at least one of the blocks. In at least some of these cases it appeared that the participants had moved their hands so that they were pressing the wrong keys and thus their responses were not recorded. Table 7 shows the descriptive statistics for the remaining participants' overall results for each difficulty level.

Table 7
Descriptive Statistics for Overall WM Task Data

Measure	Group			
	SF n = 13	SM n = 15	RF n = 14	RM n = 15
% Correct Response				
0-back	98.17 (3.66)	98.41 (3.45)	99.32 (1.73)	98.64 (2.23)[a]
1-back	94.05 (4.64)	93.81 (5.17)	95.06 (2.52)[b]	95.07 (2.72)[a]
3-back	85.12 (5.31)[c]	80.37 (7.42)	80.68 (13.92)[b]	84.80 (10.42)[b]
Correct Response RT				
0-back	497.36 (81.20)	600.23 (195.15)	527.82 (136.12)	466.03 (98.70)
1-back	612.26 (121.86)	670.20 (216.92)	656.76 (196.06)	599.56 (118.51)
3-back	851.50 (213.17)	838.50 (372.33)	853.47(249.07)	859.37 (302.02)
Incorrect Response RT				
0-back	356.83 (43.50)[d]	366.75 (80.96)[e]	402.00 (0.00)[f]	404.00 (81.19)[g]
1-back	543.38 (136.93)	737.31 (380.00)	709.24 (281.43)	674.21 (342.63)[a]
3-back	991.46 (329.65)	943.86 (493.17)	902.69 (332.88)[b]	901.68 (277.98)

Note. Mean scores are provided with standard deviations in parentheses. RTs are measures in milliseconds (ms). The n's for the incorrect response RTs on the 0-back are small because few participants made any incorrect responses on this trial.
[a]n = 14. [b]n = 13. [c]n = 12. [d]n = 3. [e]n = 2. [f]n = 1. [g]n = 5.

0-back: Correct responses. The 2 x 2 factorial ANOVA showed that there were no statistically significant main effects for Experimental Condition, $F(1, 52) = 0.80$, $p = .374$, partial $\eta^2 = .02$, or Gender, $F(1, 52) = 0.08$, $p = .779$, partial $\eta^2 < .01$, nor was there a significant interaction effect between these two variables, $F(1, 52) = 0.36$, $p = .551$, partial $\eta^2 = .01$. Thus, it appears that neither the experimental manipulation nor gender had an effect on participants' scores on this measure.

0-back: Correct response RTs. The data violated the assumptions of normality and homogeneity of variances (indicated by Levene's test). Log transformations were therefore performed and these adjustments corrected the violation of homogeneity of variances and appeared to make the data more normally distributed. A 2 x 2 factorial ANOVA showed that there were not statistically significant main effects for either Experimental Condition, $F(1, 53) = 2.02$, $p = .161$, partial $\eta^2 = .04$, or Gender, $F(1, 53) = 0.04$, $p = .845$, partial $\eta^2 < .01$. However, there was a statistically significant Experimental Condition x Gender interaction, $F(1, 53) = 4.51$, $p = .038$, partial $\eta^2 = .08$. Because statistically significant differences were not expected between the groups at this stage, post-hoc testing, using Tukey's Honest Significance Difference test, was performed on the data. This analysis found that there was a statistically significant difference ($p = .046$) between the RTs for the SM and the RM groups, with the SM participants performing more slowly than the RM participants.

0-back: Incorrect response RTs. These data were not analysed because there were so few participants with incorrect responses at this difficulty level.

1-back: Correct responses. The data appeared to be slightly in violation of the assumption of normality, but because ANOVA is a robust statistical test, this should not have had a large effect on the data. The 2 x 2 factorial analysis showed no statistically significant main effects for Experimental Condition, $F(1, 51) = 1.12$, $p = .296$, partial $\eta^2 = .02$, or Gender, $F(1, 51) = 0.01$, $p = .917$, partial $\eta^2 < .01$. There was also no significant interaction effect between these two variables, $F(1, 51) = 0.01$, $p = .907$, partial $\eta^2 < .01$.

1-back: Correct response RTs. Due to violations of the assumptions of normality and homogeneity of variances (indicated by Levene's test), log transformations were performed on the data. These transformations corrected both of these violations. The analysis showed that there were no statistically significant main effects for Experimental Condition, $F(1, 53) = 0.04$, $p = .845$, partial $\eta^2 < .01$, or Gender, $F(1, 53) = 0.01$, $p = .911$, partial $\eta^2 < .01$, and that there was no

statistically significant interaction effect between these two variables, $F(1, 53) = 0.95$, $p = .334$, partial $\eta^2 = .02$.

1-back: Incorrect response RTs. Due to violations of the assumption of normality, log transformations were performed on the data. These transformations were successful in making the data more normally distributed. The results of the 2 x 2 factional ANOVA showed that there were no statistically significant main effects for Experimental Condition, $F(1, 52) = 0.61$, $p = .439$, partial $\eta^2 = .01$, or for Gender, $F(1, 52) = 0.57$, $p = .453$, partial $\eta^2 = .01$. In addition, there was no statistically significant interaction effect between these two variables, $F(1, 52) = 2.13$, $p = .150$, partial $\eta^2 = .04$.

3-back: Correct responses. Once again the data for this measure violated the assumption of homogeneity of variances, and also showed some violations of normality. Log transformations performed on the data did not fix the violation of homogeneity of variances, and also appeared to make the normality violations worse. Therefore, the original data were used for this analysis with the knowledge that results need to be interpreted cautiously.

The analysis showed no statistically significant main effects for Experimental Condition, $F(1, 49) < 0.01$, $p = 1.00$, partial $\eta^2 < 0.01$, or for Gender, $F(1, 49) = 0.01$, $p = .908$, partial $\eta^2 < .01$. In addition, there was no statistically significant interaction effect between these two variables, $F(1, 49) = 2.70$, $p = .107$, partial $\eta^2 = .05$.

Based on the fact that the above results, especially for the main effects, are so far away from being statistically significant, it seems unlikely that correcting these violations would have moved these results much towards significance. Thus, it appears that no between-group differences exist in this dataset.

3-back: Correct response RTs. The analysis of these data did not show statistically significant main effects for either Experimental Condition, $F(1, 53) = 0.02$, $p = .884$, partial $\eta^2 < .01$, or for Gender, $F(1, 53) < 0.01$, $p = .964$, partial $\eta^2 < .01$. In addition, there was no statistically significant interaction effect between these two variables, $F(1, 53) = 0.02$, $p = .904$, partial $\eta^2 < .01$.

3-back: Incorrect response RTs. The analysis showed no statistically significant main effects for either Experimental Condition, $F(1, 52) = 0.44$, $p = .512$, partial $\eta^2 = .01$, or for gender, $F(1, 52) = 0.60$, $p = .807$, partial $\eta^2 < .01$. In addition, there was no statistically

significant interaction effect between these two variables, $F(1, 52) = 0.06$, $p = .815$, partial $\eta^2 < .01$.

Day 2: 2 x 2 x 2 x 4 ANOVAs. Descriptive statistics for the data used in the 2 x 2 x 2 x 4 (Experimental Condition x Gender x Task Difficulty x Block) ANOVAs (i.e., divided by block) are given in Table 8.

Table 8

Descriptive Statistics for WM Percentage Correct Response and Average RTs by Block

Measure		Group			
		SF $n = 12/13$[a]	SM $n = 15$	RF $n = 13/14$[b]	RM $n = 12/15$[c]
Block 1:					
CR	1-back	96.03 (3.42)	95.56 (6.61)	94.87 (5.31)	96.83 (3.10)
	3-back	87.70 (8.49)	80.63 (6.60)	80.95 (17.17)	89.68 (7.82)
RT	1-back	578.71 (138.83)	661.77 (206.58)	626.24 (209.32)	590.38 (122.07)
	3-back	909.56 (315.54)	880.75 (385.47)	886.14 (304.18)	821.72 (323.66)
Block 2:					
CR	1-back	94.05 (5.42)	94.29 (4.83)	95.61 (4.94)	96.43 (3.59)
	3-back	81.35 (8.73)	75.24 (10.83)	83.88 (13.54)	83.33 (12.27)
RT	1-back	644.59 (163.92)	686.71 (224.99)	673.19 (199.82)	565.31 (109.59)
	3-back	887.82 (249.37)	866.26 (392.89)	896.21 (273.28)	886.97 (280.44)
Block 3:					
CR	1-back	95.64 (4.29)	92.70 (8.02)	95.24 (3.87)	94.05 (4.60)
	3-back	88.49 (9.83)	81.90 (11.83)	83.52 (12.54)	90.48 (8.62)
RT	1-back	614.35 (121.49)	658.29 (215.62)	696.76 (232.37)	619.21 (143.22)
	3-back	787.89 (230.33)	757.43 (378.03)	827.24 (300.29)	839.91 (323.44)
Block 4:					
CR	1-back	94.84 (5.16)	93.33 (8.59)	94.51 (6.10)	95.64 (4.29)
	3-back	82.94 (8.72)	83.49 (10.48)	74.36 (22.38)	83.34 (15.26)
RT	1-back	611.45 (93.89)	674.05 (252.75)	630.85 (177.35)	618.59 (162.42)
	3-back	869.64 (277.27)	829.86 (362.54)	804.31 (210.17)	888.67 (336.16)

Note. Mean scores are provided with standard deviations in parentheses. CR = Percentage correct response, RT = Average correct response reaction times. Average reaction times are measured in milliseconds (ms).

[a]Data for percentage CRs based on 12 participants, data for average RTs based on 13 participants. [b]Data for percentage CRs based on 13 participants, data for average RTs based on 14 participants. [c]Data for percentage CRs based on 12 participants, data for average RTs based on 15 participants.

Correct responses. The data violated the assumptions of normality and homogeneity of variances (indicated by Levene's test) for some of the blocks; and they also violated the assumption of sphericity (indicated by Mauchley's test) for the Task Difficulty x Block interaction. Performing log transformations on the data did not correct any of these violations. Therefore, the violations of normality and homogeneity of variances were not corrected and the analysis was run with the knowledge that results would need to be interpreted cautiously. However, in order to correct for the sphericity violation, $\chi^2(5) = 11.51$, $p = .042$, I performed a Greenhouse-Geisser degrees of freedom correction ($\varepsilon = .87$) on the appropriate interaction. The results of the analysis are shown in Table 9. The scores for each experimental group for each difficulty condition are depicted graphically in Figures 9 and 10.

As expected, there was a significant main effect for Task Difficulty, accompanied by a large effect size, with participants obtaining a higher overall percentage of correct responses on the 1-back condition ($M \pm$ standard error: $M = 94.98 \pm 0.97$) than on the 3-back condition ($M \pm$ standard error: $M = 83.21 \pm 2.59$). In addition, there was a statistically significant main effect for Block and a statistically significant interaction effect for Task Difficulty and Block. These latter two results are not very informative however, as the least square means show that there was not a steady increase or a steady decrease across the combined task difficulty blocks, but rather that performances were best on blocks one and three and worse on blocks two and four. In addition, the Task Difficulty x Block least squared means indicate that the greatest differences were between the 1-back blocks and the 3-back blocks, a result already seen in the Task Difficulty main effect result. The fact that the data violated some assumptions means that it is possible that some of the results (e.g., the Experimental Condition x Gender x Task Difficulty interaction) may have reached statistical significance if these were corrected for, but because ANOVA is a robust test it therefore seems most likely that the violations would not have had a major effect on the data.

However, the most interesting statistically significant result, in terms of theory and predictions underlying the study, is that of the Experimental Condition x Gender interaction effect, which has a relatively large effect size (in comparison with many of the other effect sizes found in the WM analyses) associated with it. The interaction plots (Figure 11 and Figure 12) show that the largest differences exist between participants in the SM and RM groups. Interestingly, it can also be seen that, overall, participants in the RM ($M = 91.22 \pm 1.54$) and SF

(M = 90.13 ± 1.54) groups seemed to perform at a similar level on the WM task, and that participants the RF (M = 87.87 ± 1.48) and the SM (M = 87.14 ± 1.38) groups appeared to perform at a similar level. Thus, one tentative conclusion at this stage is that it appears that stress enhanced the performance of the female participants, while impairing performance in the male participants.

Table 9

Results for 2 x 2 x 2 x 4 ANOVA: Correct Responses

Effect	*df*	*F*	*P*	Partial η^2
Experimental Condition	1, 48	0.37	.544	.01
Gender	1, 48	0.02	.902	< .01
Experimental Condition x Gender	1, 48	4.56	.038*	.09
Task Difficulty (TD)	1, 48	86.15	< .001***	.64
Experimental Condition x TD	1, 48	< 0.01	.958	< .01
Gender x TD	1, 48	0.11	.736	< .01
Experimental Condition x Gender x TD	1, 48	3.14	.083	.06
Block	3, 144	3.13	.028*	.06
Experimental Condition x Block	3, 144	1.98	.120	.04
Gender x Block	3, 144	1.20	.310	.03
Experimental Condition x Gender x Block	3, 144	0.73	.537	.02
TD x Block	2.61, 125.48	3.93	.014*	.08
Experimental Condition x TD x Block	2.61, 125.48	2.01	.124	.04
Gender x TD x Block	2.61, 125.48	1.92	.138	.04
Experimental Condition x Gender x TD x Block	2.61, 125.48	0.62	.583	.01

* $p < .05$. *** $p < .001$.

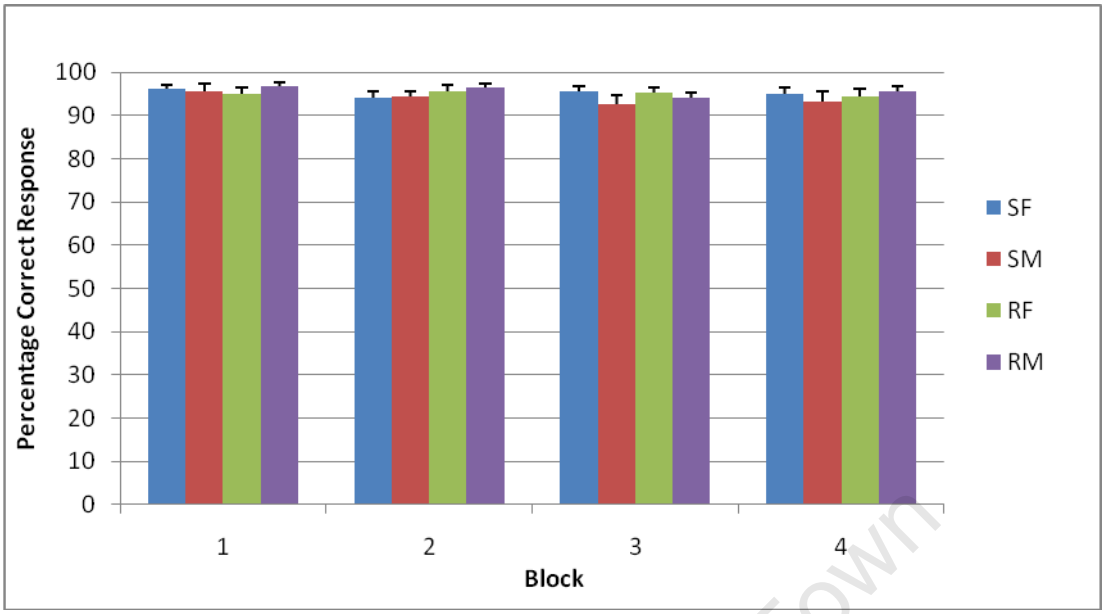

Figure 9. Percentage correct responses for 1-back condition on Day 2. Error bars indicate standard error of means.

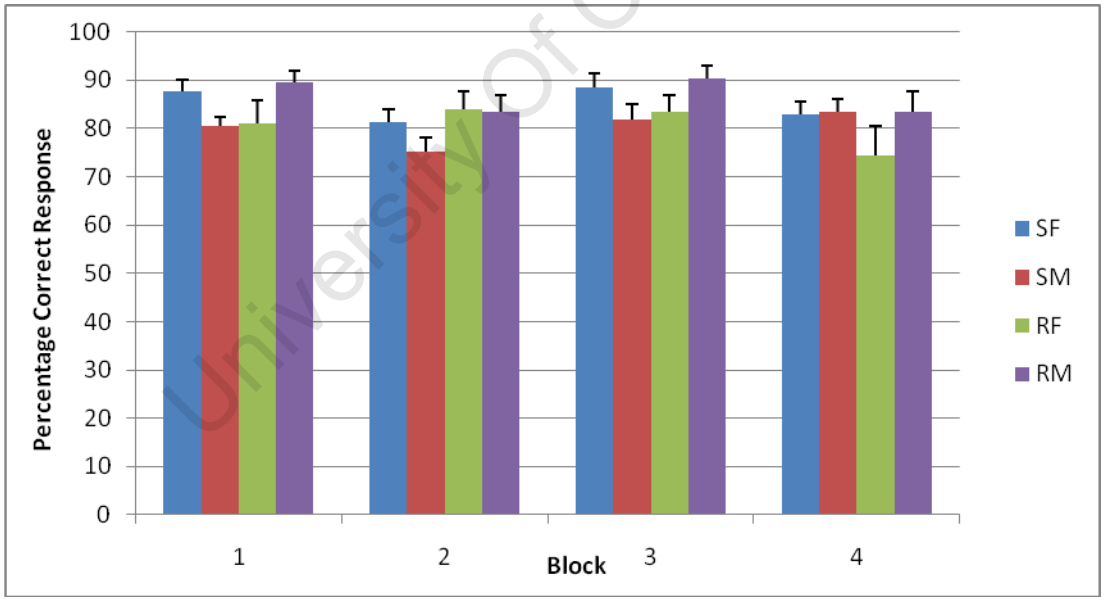

Figure 10. Percentage correct responses for 3-back condition Day 2.The error bars indicate standard error of means.

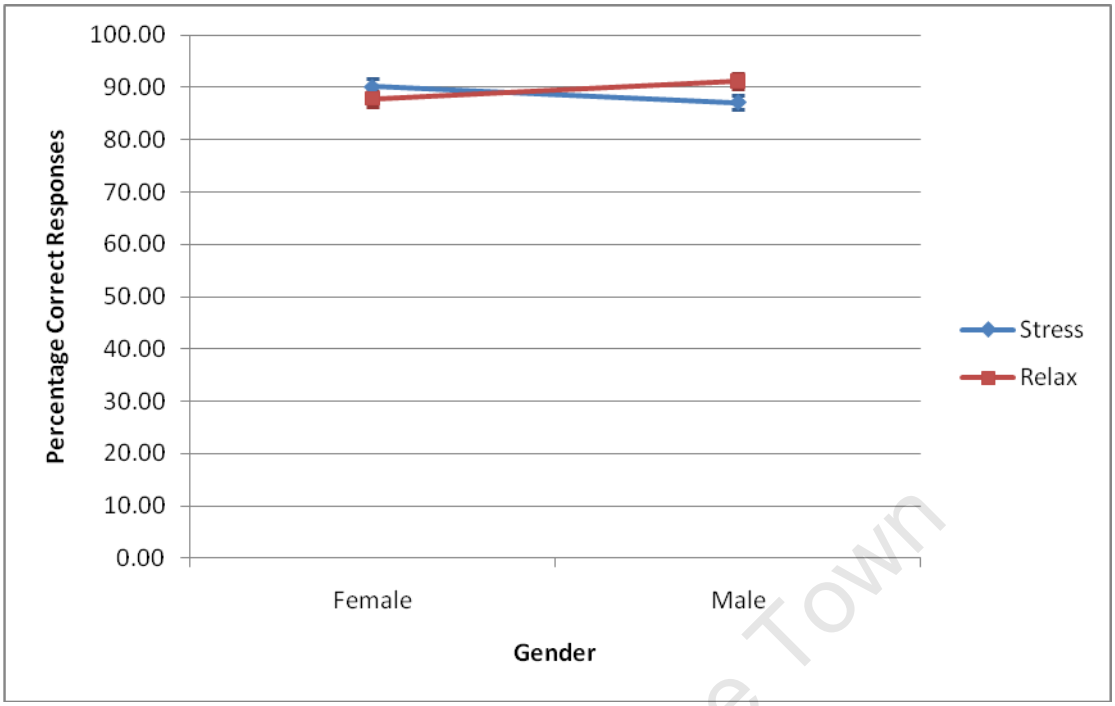

Figure 11. Interaction plot for Experimental Condition x Gender interaction. Error bars represent standard error of the mean.

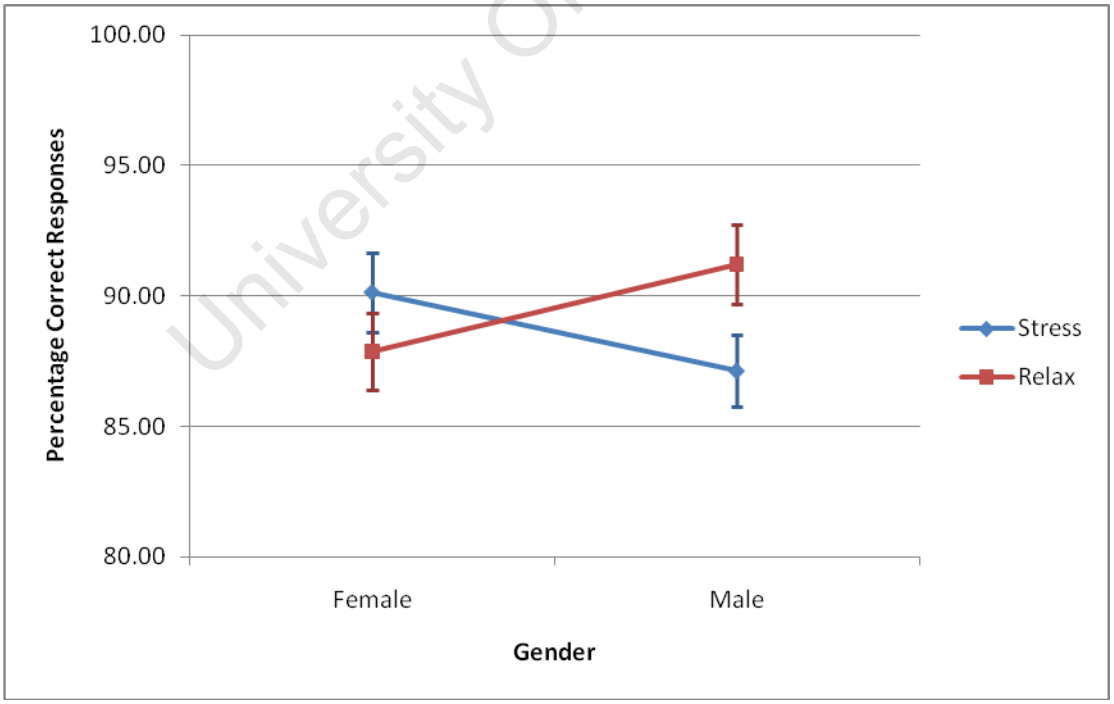

Figure 12. Enlarged interaction plot for Experimental Condition x Gender interaction. Error bars represent standard error of the mean.

Reaction times. The reaction time data violated the assumption of normality as well as the assumption of homogeneity of variances on two of the 1-back blocks. In addition, the assumption of sphericity was violated for the Block between-groups variable. Log transformations corrected the violations of normality and left only one violation of homogeneity of variances, but because I did not expect there to be differences between the groups on the 1-back condition, and because ANOVA is robust to violations, this violation should not have had a large effect on the analysis. The violation of sphericity, $\chi^2(5) = 11.86$, $p = .037$, was corrected using the Greenhouse-Geisser degrees of freedom correction ($\varepsilon = .86$) on the relevant variable.

The analysis found a statistically significant main effect for Task Difficulty which was associated with quite a large effect size, and a statistically significant Task Difficulty x Block interaction effect, which was associated with a much smaller effect size (see Table 10). Both of these results indicate that across all four blocks, overall performance on the 3-back condition ($M = 852.53 \pm 78.22$) was slower than that on the 1-back condition ($M = 634.40 \pm 45.01$).

It is clear at this point that performance on this WM task was at least partially determined by the difficulty level (indicated by both the statistically significant results and the large effect sizes on both analyses), with participants performing both more accurately and more quickly on the 1-back blocks than on the 3-back blocks. In addition, it appeared that in terms of accuracy, the experimental manipulation had more of a negative impact on the participants in the SM group, while possibly enhancing the performance of the participants in the SF group (see Figure 13, Figure 14, Figure 15 and Figure 16).

Table 10

Results for 2 x 2 x 2 x 4 ANOVA: Reaction Times

Effect	*df*	*F*	*P*	Partial η^2
Experimental Condition	1, 53	0.01	.937	< .01
Gender	1, 53	0.12	.728	< .01
Experimental Condition x Gender	1, 53	0.03	.853	< .01
Task Difficulty (TD)	1, 53	93.89	< .001***	.64
Experimental Condition x TD	1, 53	0.63	.431	.01
Gender x TD	1, 53	0.43	.517	.01
Experimental Condition x Gender x TD	1, 53	3.76	.058	.07
Block	2.58, 136.78	2.65	.051	.05
Experimental Condition x Block	2.58, 136.78	2.40	.070	.04
Gender x Block	2.58, 136.78	0.79	.501	.02
Experimental Condition x Gender x Block	2.58, 136.78	1.45	.229	.03
TD x Block	3, 159	8.15	< .001***	.13
Experimental Condition x TD x Block	3, 159	1.51	.213	.03
Gender x TD x Block	3, 159	1.64	.183	.03
Experimental Condition x Gender x TD x Block	3, 159	0.53	.666	.01

***$p < .001$.

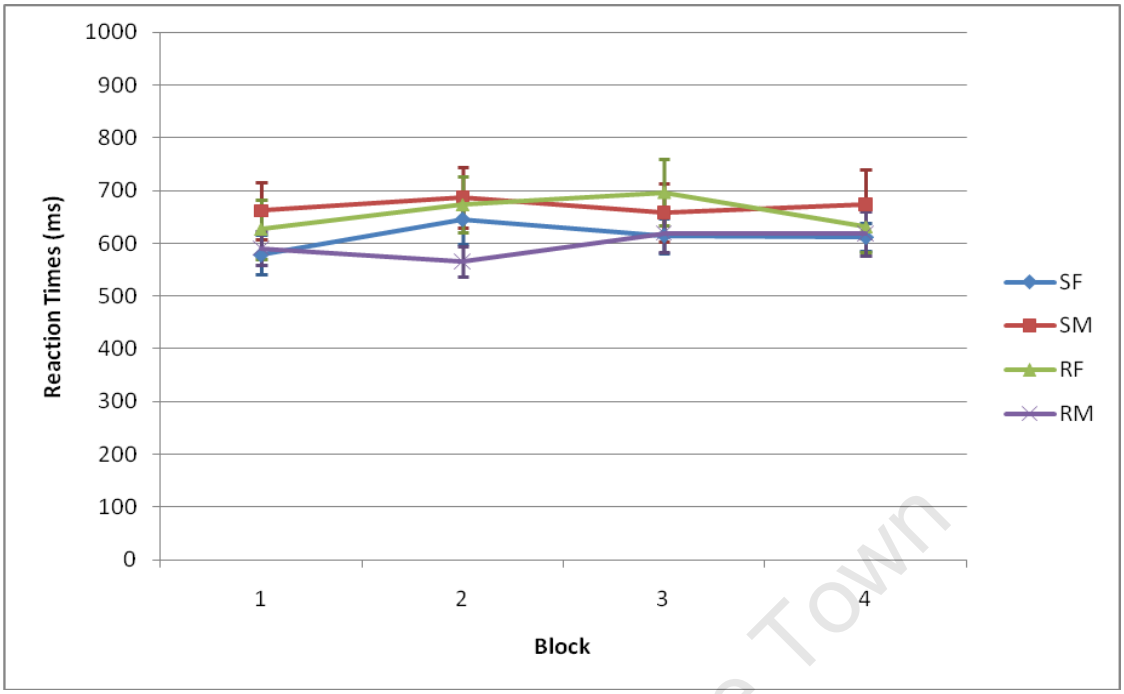

Figure 13. Reaction times for correct responses: 1-back condition on Day 2. Error bars indicate standard error of the mean

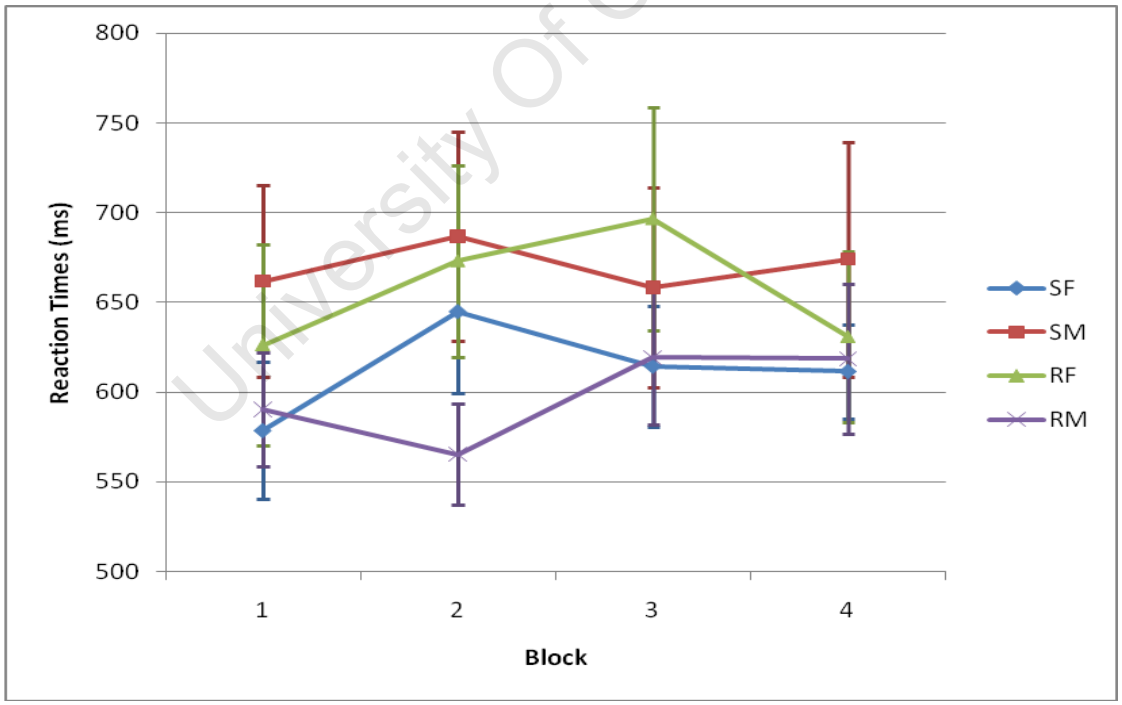

Figure 14. Englarged depiction of reaction times for correct responses: 1-back condition on Day 2. Error bars indicate standard error of the mean

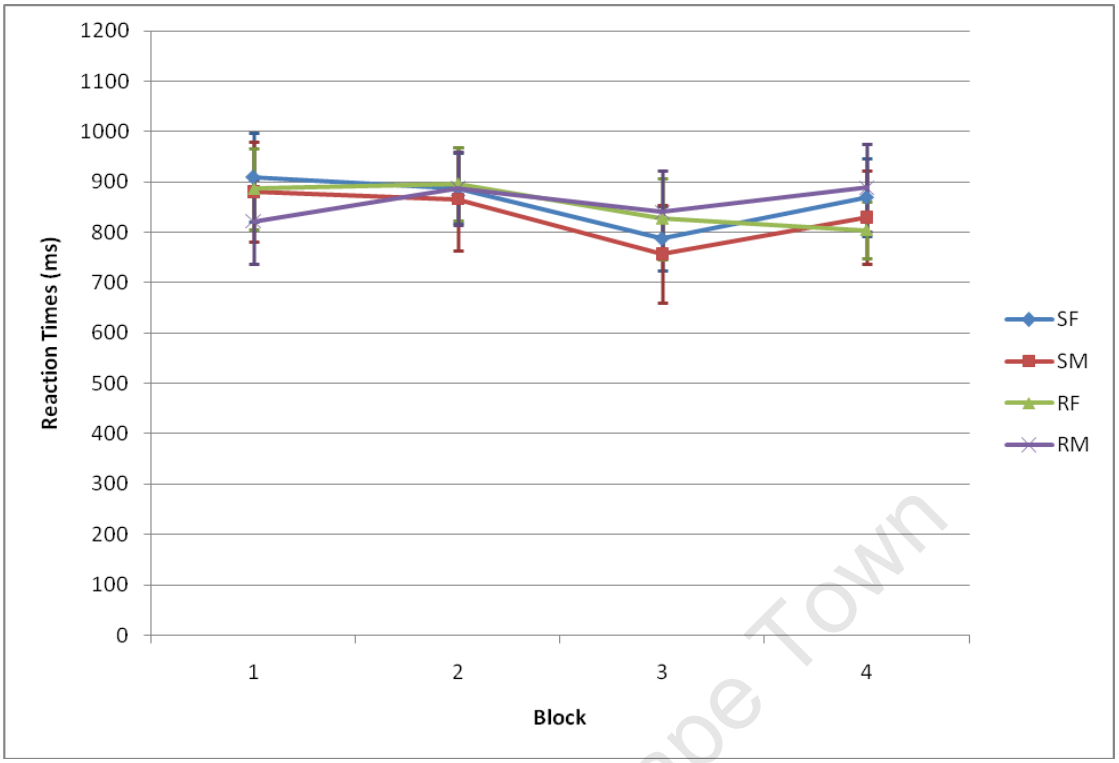

Figure 15. Reaction times for correct responses: 3-back condition on Day 2. Error bars indicate standard error of the mean.

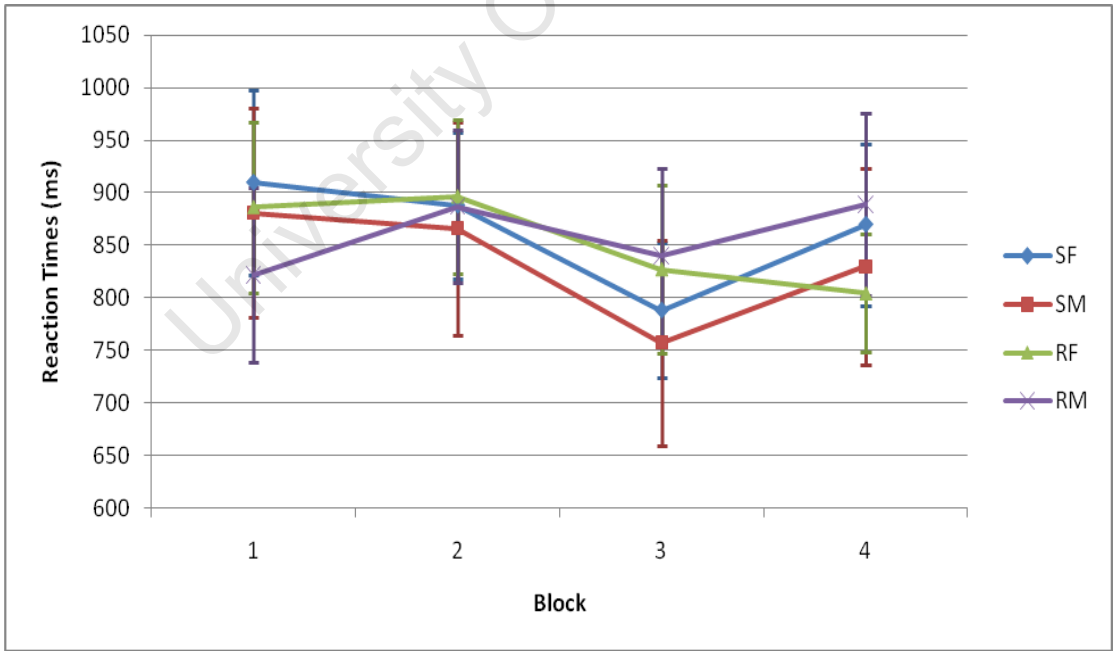

Figure 16. Enlarged depiction of reaction times for correct responses: 3-back condition on Day 2. Error bars indicate standard error of the mean.

Day 2: Males only - 2 x 2 x 4 ANOVA. Based on the observations from the 2 x 2 x 2 x 4 repeated measures ANOVA on the correct responses data, and from my observations of the descriptive statistics (Table 8), I decided to run an identical analysis to Schoofs et al. (2008) and include only the male participants from this study, as it appeared that it was between the SM and RM groups that the greatest differences in WM performance lay. Thus, I ran a 2 x 2 x 4 repeated measures ANOVA on their WM data (both percentage of correct responses and correct response RTs).

Correct responses. Although there were slight deviations from normality on some of the sets, ANOVA is a robust statistical test and thus these deviations should not have affected the results. The analysis showed a similar pattern of results to that reported by Schoofs et al. (2008), including statistically significant main effects for Experimental Condition and for Task Difficulty (the former being associated with a medium affect size and the latter with a large effect size). In addition, there was a statistically significant interaction effect between Task Difficulty and Block (see Table 11 for complete results). As expected, and as can be seen from the descriptive statistics presented in Table 8, and from the graphical depictions in Figure 17, the greatest differences exist on the 3-back condition. Therefore, I performed planned contrasts on the four 3-back condition blocks.

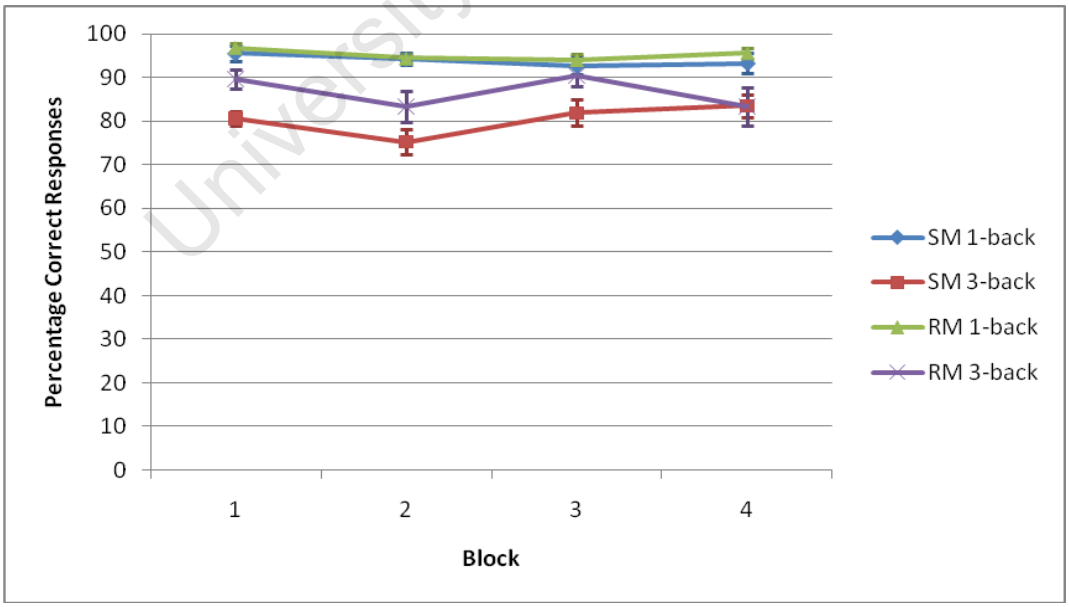

Figure 17. Percentage of correct responses for the male participants across the 1-back and 3-back conditions on Day 2. Error bars indicate standard error of the mean.

These contrasts showed statistically significant results for the first, $F(1, 25) = 10.64$, $p = .003$, and third blocks, $F(1, 25) = 4.41$, $p = .046$, and a close to statistically significant result for the second block, $F(1, 25) = 3.32$, $p = .081$. The fourth block's contrast was not statistically significant, $F(1, 25) < 0.01$, $p = .975$.

Table 11
Results for 2 x 2 x 4 ANOVA: Correct Responses

Effect	*df*	*F*	*p*	Partial η^2
Experimental Condition	1, 25	4.53	.043*	.154
Task Difficulty	1, 25	59.97	< .001***	.71
Experimental Condition x Task Difficulty	1, 25	2.49	.127	.09
Block	3, 75	1.85	.146	.07
Experimental Condition x Block	3, 75	0.91	.439	.04
Task Difficulty x Block	3, 75	4.65	.005**	.16
Experimental Condition x Task Difficulty x Block	3, 75	2.00	.121	.07

*$p < .05$. **$p < .01$. ***$p < .001$.

Reaction times. Due to violations of the normality assumption, the homogeneity of variances assumption on two of the 1-back sets, and sphericity, log transformations were performed on the data. The transformations, while making the data more normally distributed, did not correct for the violations in homogeneity of variances (although they did improve the results) and only corrected one of the sphericity violations. For the Block sphericity violations, $\chi^2(5) = 14.56$, $p = .013$, the Greenhouse-Geisser correction for degrees of freedom ($\varepsilon = .76$) was applied.

The analysis showed a significant main effect for Task Difficulty (once again associated with a large effect size) and significant interaction effects for Task Difficulty and Block as well as Experimental Condition and Block. However, there was no significant main effect for Experimental Condition (see Table 12 for complete results). As can be seen by the descriptive statistics in Table 8, and the graphical depiction in Figure 18, the Stress group was performing slower (although with less accuracy) on the 1-back condition and faster (although again with less accuracy) on the 3-back condition, than the Relax group was. However, because there was no

statistically significant main effect for Experimental Condition, and because the larger ANOVA did not indicate anything of particular interest, further analyses were not performed on these data.

The results from the above two analyses indicate that the male participants in the Stress group may have been negatively affected, particularly in terms of accuracy, on the more difficult condition of the WM task. It does not appear, however, that their RTs were significantly affected. Thus, unlike the Schoofs et al. (2008) study, this study suggests that only one aspect of the WM task was affected by increased cortisol levels.

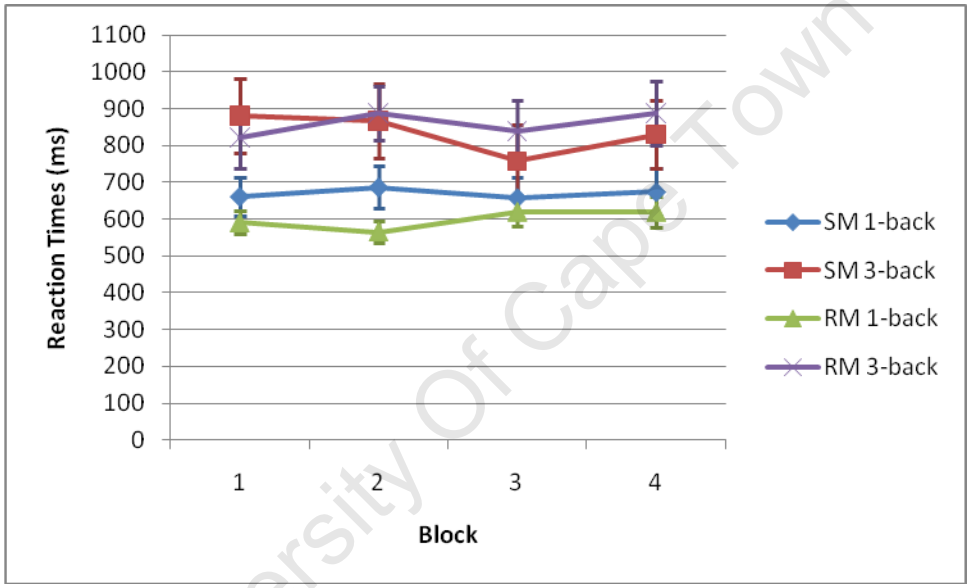

Figure18. Reaction times for correct responses for the male participants across the 1-back and 3-back conditions on Day 2. Error bars indicate standard error of the mean.

Table 12

Results for 2 x 2 x 4 ANOVA: Reaction Times

Effect	*df*	*F*	*p*	Partial η^2
Experimental Condition	1, 28	< 0.01	.945	< .01
Task Difficulty	1, 28	33.47	< .001***	.55
Experimental Condition x Task Difficulty	1, 28	3.06	.091	.10
Block	2.28, 63.75	1.38	.259	.05
Experimental Condition x Block	2.28, 63.75	3.18	.042*	.10
Task Difficulty x Block	3, 84	4.02	.010*	.13
Experimental Condition x Task Difficulty x Block	3, 84	2.43	.071	.08

*$p < .05$, **$p < .01$, ***$p < .001$

Discussion

Summary and Implications of Results

Experimental manipulation. From the analysis of the self-report and physiological data for those participants included in the final sample, it is clear that the experimental manipulation worked as predicted. The stress-induction procedure induced a physiological stress response, significantly increasing cortisol levels, whereas the relaxation procedure significantly decreased cortisol levels. The cardiovascular and self-report anxiety data were congruent with the cortisol data. Thus, one can conclude that, as planned, at the start of the Day 2 memory testing phase participants in the two Stress and the two Relax groups were experiencing different physiological and psychological states. This fact further implies that these groups were experiencing differential activation of the hippocampal and PFC regions. My hypothesis was that these different cortisol levels would lead to relative impairments in hippocampal- and frontal-based memory functioning (i.e., impaired delayed cued recall DM and impaired WM, respectively) in the participants in the two Stress groups.

Declarative memory. This study investigated the effects of stress on DM, specifically the retrieval of the delayed cued recall and recognition of previously-learnt material.

Cued recall.

Day 1. As expected, the participants in all groups, regardless of gender, all recalled a greater number of words on the second cued recall trial than on the first, indicating that the second trial was beneficial to the encoding process. Interestingly though, and in comparison to the versions of this test used in the WMS-III (Psychological Corporation, 1998) and by Uttl et al. (2002), the version of the test used in the current study appears to be a more efficacious test of cued recall DM.

In looking at the total number of word pairs recalled by participants, of the 85 participants who completed the study, 62 (72.94%) achieved or surpassed the maximum possible score on the WMS-III VPA subtest (i.e., 8 word pairs) on the first immediate cued recall trial, with only two participants (2.35%) not achieving this score on the second immediate cued recall trial. In addition, 16 participants (18.82%) achieved or surpassed the VPA-15's (Uttl et al., 2002) maximum possible score on the first immediate cued recall trail, with almost half the participants (49.41%) achieving or surpassing this score on the second immediate cued recall trial. In contrast, on the version of the test used in the current study, no participant achieved a perfect

score on the first immediate cued recall trial, and only 17 (20%) achieved a perfect score on the second cued recall trial. In other words, it appears that where ceiling effects might be problematic for the WMS-III VPA and VPA-15 tasks, the VPA test used in this study had no such problems.

It is important to note, however, that this study's VPA test included three more easy word pairs than the VPA-15, and that the WMS-III does not include any easy pairs. Thus, I contrasted just the scores for the difficult pairs from my test with those of the other two. Of the 85 participants who completed the study, 17 (20%) recalled eight or more difficult word pairs on the first immediate cued recall trial, while just over half (52.94%) of the participants recalled eight or more difficult word pairs on the second immediate cued recall trial. When using the VPA-15 as a comparison, which had the same 11 difficult pairs as my study, no participants in my study achieved the recall of 11 difficult pairs on the first immediate cued recall trial, but 17 (20%) achieved the maximum possible difficult pair score on the second immediate cued recall trial.

Thus, even with the added easy word pairs on my task, more participants reached the maximum score (18) on the second cued recall trial than reached the maximum score in Uttl et al.'s (2002) study (15; 20% vs. 14.5 %). It therefore appears that the version of the task used here is definitely a more efficacious measure of hippocampal-dependent declarative memory than the WMS-III task, as only a small percentage of participants attained ceiling scores on my task, compared to the WMS-III where massive ceiling effects are shown (Uttl et al., 2002).

Day 2. I tested the hypothesis that cued recall memory would be affected by stress, especially among male participants. Although the results of the data analyses showed that the number of word pairs recalled on Day 2 was lower than that on Day 1, they also showed that neither the experimental manipulation nor gender significantly affected delayed cued recall performance. These results are in contrast to those of de Quervain et al. (2003) and of Kirschbaum et al. (1996), who reported that increased cortisol levels caused impairment in delayed cued recall performance. My results are consistent, however, with those of Kuhlmann et al. (2005) and of Lupien et al. (1999), who reported that delayed cued recall performance was not affected by increased cortisol levels.

It is important to note, however, that the protocols of the Kirschbaum et al. (1996, Study 1 and 2) and Lupien et al. (1999) studies were quite different to mine, whereas the protocols of the de Quervain et al. (2003) and Kuhlmann et al. (2005) studies were quite similar to mine. The

two Kirschbaum et al. (1996) studies and the Lupien et al. (1999) study (a) raised cortisol levels prior to the encoding of the material to be recalled, and (b) only investigated between 5- and 30-minutes delayed cued recall. The de Quervain et al. (2003) and Kuhlmann et al. (2005) studies both (a) increased cortisol levels just before retrieval, (b) investigated 24-hour delayed cued recall, and (c) used intentional encoding of their declarative memory material. Thus, it is the results of these two studies latter studies which are of most comparative interest here.

Interestingly, the study that artificially increased cortisol levels (using cortisone; de Quervain et al., 2003) found impaired delayed cued recall memory performance, whereas the study that used naturalistic means to raise cortisol levels (using the TSST; Kuhlmann et al., 2005) did not find impaired delayed cued recall memory performance. The differences in participants' cortisol levels between these two studies are marked, with the first study reporting mean cortisol levels of 50.3 nmol/l at the time of memory testing and the second study reporting mean levels of only about 17 nmol/l at the time of testing. These latter levels are obviously much closer to the levels achieved in my study. Thus, the results (viz., impaired delayed cued recall performance) seen in the de Quervain et al. (2003) study could have been achieved as a result of the much higher cortisol levels, which were not achieved by means of psychosocial stress in either the Kuhlmann et al. (2005) study or the current study. It is possible, therefore, the procedures used by de Quervain et al. (2003) increased cortisol levels to the extent that they existed on the right hand side of the inverted-U curve (thus impairing delayed cued recall performance), while those used by Kuhlmann et al. (2005) and by the current study did not increase cortisol levels sufficiently to either negatively or positively affect this type of memory functioning.

It is also interesting to note that the Kuhlmann et al. (2005) study was conducted during the late morning, whereas the current study was conducted during the late afternoon and early evening. Because the two studies achieved similar results across different times of day, it appears that time of day may not significantly moderate the effect of stress on delayed cued recall declarative memory.

In summary, the data from the current study are consistent with those of some (but not all) previous studies in showing that delayed cued recall does not appear to be affected by stress experienced during retrieval of previously-learnt material. However, because there have been so many conflicting results in this area, further studies with more extensive controls are needed in

order to fully elucidate this area. Possible designs for these studies are outlined later in this paper.

Recognition. With regards to recognition, I tested the hypothesis that (a) performance would remain unaffected by increased cortisol levels, and (b) the participants in the Stress Female group would perform better than those in the Stress Male group. In terms of the number of correct identifications (i.e., hits) made by the participants on the recognition task, it did not appear that either the experimental manipulation or gender significantly affected performance. It does, however, appear from Table 6 that participants in the Stress Female group did perform slightly better than participants in any of the other three groups.

Although this set of results only partially confirms the *a priori* hypotheses, they are consistent with the basic results of the other studies that have increased cortisol levels at the retrieval phase of the recognition memory process (Buchanan & Tranel, 2008; de Quervain et al., 2000, 2003; Domes, Heinrichs, Rimmele, Reichwald, & Hautzinger, 2004; Lupien et al., 2002, Study 2). The results of the current study also support Lupien et al.'s (1999) hypothesis that stress induced at the retrieval stage would not affect recognition performance. Although not strictly relevant to the current study, it is interesting to note that the other part to the Lupien et al. (1999) hypothesis was that recognition would be affected if the memory process was disrupted by stress at the encoding phase. However, many studies have disconfirmed this hypothesis, finding that increased cortisol levels at encoding also do not impair recognition performance (de Quervain et al., 2003, Domes et al., 2004, Luethi et al., 2009; Nater et al., 2007; Tops et al., 2003). Together, these results suggest that recognition performance is not affected by stress experienced at any point of the memory process.

There is evidence, however, that while overall recognition performance and recognition for *neutral* material (e.g., of the type used in this study) is not affected by increased cortisol levels, these increases may affect recognition for *emotional* material (e.g., positively or negatively valenced words; Domes et al., 2004; Tops et al., 2003). Thus, although the results of the current study do not show impairments in recognition performance as a consequence of stress, it is possible that these data were only obtained because emotionally neutral material was used as test stimuli.

Although there were no differences between the participants in terms of accuracy on the recognition measure, I did find that there was a statistically significant difference in reaction

times between the genders for their correct responses. Specifically, male participants appeared to have significantly slower reaction times when making correct responses than female participants did. This is an interesting result that is not a consequence of the experimental manipulation (because this pattern was consistent across the Stress and Relax groups), but that could be explained by the fact that women are generally better at recognition tasks than men are (Maitland et al., 2004), and therefore may be able to complete them more quickly.

With regards to how time of day may affect the impact of stress on this memory process, because findings (for neutral material at least) are consistent across studies (which have been conducted at all different times of the day), it seems unlikely that this form of declarative memory is moderated by this factor. In addition, it also therefore does not appear that recognition memory for neutral material is moderated by cortisol levels (which have varied widely across studies); therefore performance on recognition tasks may not fit into the predictions made by the inverted-U curve hypotheses as has well as performance on other DM tasks.

The results of the current study therefore support the findings of the previous literature with regard to neutrally valenced material; namely, that the retrieval of such material is not affected by stress applied after the encoding stage of the memory process. However, because so many of the stimuli experienced in real life are emotionally charged, it is essential that emotional material is further investigated in future studies.

Working memory. I tested the hypothesis that the 3-back condition of my WM task would be more affected by stress than the 1-back condition. I also tested the hypothesis that participants in the Stress Female group would perform more accurately, but more slowly, than participants in the Stress Male group.

As expected, the analyses of the *n*-back data showed that performance on the 3-back condition was generally less accurate and slower than for 1-back condition, regardless of the experimental group or gender of the participants. In addition, and again as expected, stress did not appear to affect performance on the 1-back condition. Against expectations, however, stress also did not appear to significantly affect participants' reaction times on any of the conditions, nor did the decreased performance on the 3-back become universally more pronounced for the participants in the Stress group. Specifically, although stress did have a significant, negative

effect on the accuracy of participants in the Stress Male group, it actually appeared to enhance the accuracy of participants in the Stress Female group.

It is probably that so many previous studies in this area have found such consistent impairments in working memory performance under stress because they only included male participants (e.g., Luethi et al., 2009; Lupien et al., 1999; Oei et al., 2006; Schoofs et al., 2008, Wolf, Convit, et al., 2001). The results from the current study indicate that had female participants been included in previous research, results from these studies may not have been as definitive as they currently appear. Although Elzinga and Roelofs (2005), as previously noted, did include female participants in their sample, their final sample size was too small to draw any definite conclusions about whether sex differences existed under conditions of stress. In addition, although Kuhlmann et al. (2005), as previously noted, did not find impairments in working memory following stress, they used the digit span task, which is not the most efficacious measure of working memory.

Although it does not appear that the results of the current study are consistent with previous research in this field, there is evidence that frontal lobe functioning may benefit from the experience of stress in women, but be adversely affected by the experience of stress in men. Van den Bos, Harteveld, and Stoop (2009) found that, under conditions of stress, males made more risky decisions in the Iowa Gambling Task (a test of decision making related to ventromedial PFC functioning; Bechara, Damasio, A. R., Damasio, H., & Anderson, 1994) than did non-stressed male participants. Women under stress, on the other hand, made fewer risky decisions with smaller cortisol increases and more risky decisions with greater cortisol increases. In the van den Bos et al. (2009) study, the women classified as 'low' cortisol responders had post-stress levels of about 9 nmol/l, whereas those classified as 'high' cortisol responders had post-stress levels of above 13 nmol/l. Clearly, by these standards participants in the Stress Female group of the current study would be classified as low responders. Therefore, the results of the current study are consistent with those of the van den Bos et al. (2009) study in showing that under conditions of stress where cortisol levels are not raised beyond a particular threshold, frontal lobe functioning will be enhanced in women, whereas at all cortisol levels it will be impaired in men.

Although based on the 3-back accuracy results it is understandable that there were no differences between the participants in the Stress Female and Relax Female groups for reaction

times, it is difficult to ascertain why this study did not replicate Schoofs et al.'s (2008) statistically significant differences on this variable between the participants in the Stress Male and Relax Male groups. One possible explanation is that, unlike in Schoofs et al.'s (2008) study, participants practiced the task before undergoing the experimental manipulation. This practice may have reduced any differences in performance: It has been shown that practice effects on working memory tasks can cause an improvement in performance, even under conditions of stress (Elzinga & Roelofs, 2005; Schoofs et al., 2008). Thus, it is possible that reaction time as a performance outcome variable may have been more sensitive to practice effects than accuracy, thereby reducing the differences between the participants in the Stress Male and Relax Male groups.

In terms of the specific hypotheses tested in this study regarding differences between the sexes on working memory related hypotheses of this study (i.e., that participants in the Stress Female group would perform more accurately, but more slowly, than those in the Stress Male group), although the results were not statistically significant, it was true that participants in the Stress Female group performed more accurately, but more slowly, than those in the Stress Male groups As already stated, more research into exactly why these sex differences exist is necessary as it is currently scarce in this area, and definitive conclusions cannot be drawn from this study alone.

In terms of the time of day effects of stress on working memory performance, it is not possible to draw any definitive conclusions from this study because (a) this appears to be the only study that has found a sex differences in working memory performance, and (b) most previous studies in this area were conducted during the morning hours. Therefore, it is difficult to make comparisons and say for certain whether this sex difference would exist at other times of day. However, with regard to the effects of stress on working memory performance in males, based on the findings of this study and others, it appears that these individuals are negatively affected, at least in terms of accuracy if not reactions time, at all points during the day.

Limitations and Directions for Future Research

Low baseline cortisol levels and small cortisol increases. Although the acute psychosocial stressor utilised in the current study did successfully increase cortisol levels from baseline, the baseline and post-TSST cortisol levels of the Stress Female and Stress Male

participants in the final sample were quite different to those in other studies employing this stressor. This is an important fact to note because it might have potentially affected these participants' memory performances.

The average increase in cortisol levels across all of the final sample Stress group participants was 4.91 nmol/l (SF = 3.65 ± 2.72; SM = 6.01 ± 3.86 nmol/l). This increase is at the lower end of the range of average increases reported by other studies in this field. Previous studies have reported average increases of between about 4 nmol/l and 15 nmol/l, with many reporting average increases of either over 9 nmol/l (e.g., Domes et al., 2004; Kirschbaum et al., 1996; Luethi et al., 2009; Nater et al., 2007; Oei et al., 2006; Wolf, Schommer, et al., 2001) or of between about 4 nmol/l and 5 nmol/l (e.g., Elzinga & Roelofs, 2005; Kuhlmann et al., 2005; Schoofs et al., 2008). In addition to the post-manipulation increases in the current study being relatively small (especially taking into account the fact that only the cortisol responders were included in the final sample), both the baseline and post-TSST cortisol levels were also much smaller than those reported in previous studies. Most other research in this field has reported baseline levels of over 7 nmol/l, and post-TSST levels of over 12 nmol/l (see, e.g., Domes et al., 2004; Elzinga & Roelofs, 2005; Kirschbaum et al., 1996; Kuhlmann et al., 2005; Luethi et al., 2009; Nater et al., 2007; Oei et al., 2006; Schoofs et al., 2008; Wolf, Schommer et al., 2001), whereas in the current study, average baseline levels were only 1.62 nmol/l and average post-TSST levels were only 6.53 nmol/l.

Thus, not only did the current TSST exposure result in relatively small increases in salivary cortisol levels, but the post-TSST cortisol levels were also relatively low. Following the logic of the inverted-U curve hypothesis, the levels of cortisol in this study may have had an effect on participants' performance on the memory tests (i.e., the cortisol levels may have fallen closer to the middle of the curve than in previous studies where they may have fallen further to the right, resulting in more optimal memory performance in this study than in previous studies), and therefore make comparisons between the results of the current study and previous studies difficult.

These differences in cortisol levels between the current study and previously published studies might be attributed to time of day effects. As already discussed, very few published studies in this area have been conducted in the afternoon, and none have been conducted as late in the day as the current study (i.e., between 16h00 and 20h00). However, the reported baseline

cortisol levels of participants in previous studies conducted in the afternoon have generally been greater than 8nmol/l, with post-stressor levels generally being greater than 13 nmol/l (Kirschbaum et al., 1996, Study 1; Luethi et al., 2009; Nater et al., 2007). Hence, it seems unlikely that time of day effects alone could account for the low baseline cortisol levels in the current participants. In addition, the fact that it has been noted that lower baseline cortisol levels should result in greater cortisol increase in response to a psychosocial stressor (Kudielka et al., 2004; Maheu et al., 2005), means that the current study should actually have seen greater cortisol responses to the stressor than previous studies due to the lower baseline cortisol levels of its participants. This clearly does not appear to be the case, however.

Two previous studies conducted in our laboratory (Bonito Atwood, 2008; Henry, 2008) also found lower basal cortisol levels in their participants than reported in previously published research from other laboratories (M = 5.93 ± 2.76 nmol/l and M = 3.01 ± 3.56 nmol/l respectively). As Bonito Atwood (2008) pointed out, the differences cannot be age-related as the participants in her study (and in the other studies in our laboratory) are of a similar age range to the participants in the research to which this study is being compared. In addition, although the smaller increases experienced in the previous two studies run in our laboratory (M = 2.90 nmol/l and M = 2.65 nmol/l, respectively) could possibly have been accounted for by variations in the TSST procedure from the original protocol, the current study aimed to follow the original protocol much more closely. As a result, an explanation related to the efficacy of the stressor being used is clearly not a satisfactory reason for the low cortisol levels reported in the current study, and therefore points to the likelihood that it is not a satisfactory explanation for the similar findings in our laboratory's previous TSST-based research.

Taking all of the above into account, it seems likely that there is something particular about the samples being used in the studies from our laboratory that makes their basal cortisol levels, and their degree of cortisol response to psychosocial stress, so different from those reported in previous literature. One possible explanation[3] for this is the fact that the samples tested by studies in our laboratory have, unlike previous studies, been drawn from a South

[3] Another possible explanation is that the laboratory used to analyse the cortisol samples collected from our studies may not have been conducting the analyses in the same way as in previous literature. However, the laboratory we used has previously found salivary cortisol levels similar to those reported in other research (Pillay, Haumann, Bonito Atwood, Omar, Thomas, 2008) and it therefore seems unlikely that this is a valid explanation.

African population. Given the racial heterogeneity of the South African population, and of our samples compared to previous studies, it is possible that this factor may help to provide an explanation for the discrepant cortisol levels.

Effect of race on HPA axis response. Chong, Uhart, McCaul, Johnson and Wand (2008) measured subjective anxiety and cortisol level changes in response to the TSST in a sample of American adults who were divided into race groups based on their self-report. The results showed that although White and Black participants scored similarly on measures of subjective anxiety in response to the TSST, their physiological responses differed significantly, with White participants showing greater cortisol increases than Black participants.

Although to my knowledge no other study has attempted to replicate this finding, the data reported from the Chong et al. (2008) study does suggest that there may be a racially moderated response to psychosocial stressors. This racially-based difference could, of course, have important consequences for the effects of stress on cognitive performance. It is important to note, however, that Chong et al.'s (2008) sample consisted of men and of women in the follicular phase in their cycle. In addition, the two racial groups did not have an even male:female ratio, with there being a greater percentage of women in the Black group than in the White group. Because it has been shown that women in the follicular phase of their cycle may have significantly smaller increases in cortisol in response to the TSST than men (Kirschbaum et al., 1999; Uhart et al., 2006), this unevenness in the sample distribution may prove to be a confounding variable and may partially explain why the Black participants experienced smaller cortisol increases than the White participants.

Despite this potential confounding variable, these possible effects of race on the stress response (and thus potentially on its cognitive consequences) are important to consider when designing studies in this area, particularly if those studies are to take place in a racially heterogeneous country such as South Africa. Because this factor was not taken into account in the current study, it is possible that this potential racially-based difference in HPA-axis response to the TSST may have the memory performance data.

Of the 85 participants who completed the current study's experimental protocol, 37 were considered by the researchers to be White, 26 Black, 20 Coloured or Indian, and 2 Asian. Of the participants retained in the final sample, 20 were White, 21 Black and 16 Coloured or Indian. Thus, it is clear that the sample used in this (and most other South African) research is probably

a lot more heterogeneous than the samples used in the majority of previous studies in this field. Previous studies have generally been conducted in either Europe (e.g., Germany, Netherlands of Switzerland; de Quervain et al., 2000, 2003; Domes et al., 2004; Elzinga & Roelofs, 2005; Kirschbaum et al., 1996; Kuhlmann et al., 2005; Luethi et al., 2009; Nater et al., 2007; Oei et al., 2006; Schoofs et al., 2008; Tops et al., 2003; Wolf, Schommer, et al., 2001) or North America (Lupien et al., 1999; Lupien et al. 2002; Wolf, Convit, et al., 2001). Therefore, it is most probable that these previous studies' samples consisted of mostly White participants, although it is not possible to know this definitively, because no racial information is supplied by the authors of these studies.

Unfortunately, the final sample from the current study was too small to run any meaningful analyses on the cortisol levels for the different racial groups. However, in future studies, especially in a country such as South Africa, it is important to bear this variable in mind, as it could have important consequences for understanding inter-individual differences in the effects of stress on cognitive functioning.

Effects of sex on HPA axis response.

Sex differences in the effects of stress on cognitive functioning. As previously outlined, the issue of sex differences in the effects of stress on memory performance has not been explored sufficiently. Although the current study aimed to investigate this area, and did find some interesting results, it is still necessary for future studies to focus more closely on this subject.

Future research should aim to include females in different phases of their menstrual cycles, as well as those that are on oral contraception, in their samples. Although the reasons for controlling these variables are very understandable (viz., so that HPA axis responses, and thus cognitive performance differences, between men and women can be reliably compared), responses to stress and the resulting cognitive performance during the late luteal phase of the menstrual cycle is clearly not a generalizable state for all women all of the time. Obviously, women will experience stress during all stages of their menstrual cycle, not only during the late luteal phase of their cycle. Therefore, it is necessary to also investigate more thoroughly how female cognitive performance is affected by stress at all stages of the menstrual cycle. In addition, a large number of women use some form of oral contraceptive, and therefore their functioning under conditions of stress is also a relevant question. It is also important to

investigate how women in the different phases of the menstrual cycle, or using oral contraception, differ from men in their cognitive performance under stress. Findings related to this question would have real-world implications for sex differences in cognitive functioning, especially seeing as it has been reported that cognitive performance among females may fluctuate over the different stages of the menstrual cycle (Kimura & Hampson, 1994). Thus, further studies including a diverse sample of young female participants are needed in this field in order to provide a more comprehensive view of how stress affects individual cognitive functioning.

Effects of the type of psychosocial stressor employed on HPA axis functioning in men and women. Although the recorded cortisol levels in the current study are much lower than those reported in previous research, participants in the Stress Male group had much greater cortisol responses to the stressor and had higher post-TSST levels than participants in the Stress Female group. This finding is consistent with previous literature (i.e., other studies have also found the same trend of cortisol level increase in response to stress among males than females). In addition, as is reflected in the current study, this difference appears to exist even when phase of menstrual cycle and oral contraceptive use are controlled for (Elzinga & Roelofs, 2005; Kirschbaum et al., 1999; Wolf, Schommer, et al., 2001).

Although it is generally thought that these differential responses by men and women to the TSST reflect universal sex differences in HPA-axis responses to psychosocial stress, it has been suggested that men and women may have different HPA-axis responses to different types of stressors (Gillespie & Eisler, 1992).

Stroud et al. (2002) provided empirical evidence to support these theories. Specifically, they described statistically significantly different patterns in cortisol response between men and women dependent on the type of stressor to which they were exposed. In response to what they classified as an 'achievement stressor', which consisted of both an evaluated verbal challenge and an evaluated mathematical challenge, men showed much greater cortisol increases than women, who actually appeared to show a slight decrease in cortisol levels (a similar result to that of the current study). However, in response to what they classified as a 'social rejection challenge', which involved participants engaging in social conversations and then being slowly ostracised from these, women showed much greater cortisol increases than men. In fact, in the latter task, women showed approximately six times the cortisol increases that they showed in the

former task. In contrast, men showed 145 times the increase in cortisol levels in the former task compared to the latter. Thus, it appears that women may in fact have an HPA-axis response to a stressor that is the same as or greater than men, depending on what exactly the stressor is that they are experiencing. Because under the above classification the TSST would be considered to be an achievement stressor, it should affect men to a greater extent than it should women; a fact that would account for the differential cortisol increases found between the sexes in response to the TSST.

Therefore, it is clear that future studies in this field may need to involve different types of psychosocial stressors in order to investigate whether these will differentially affect cognitive performance between the sexes. These studies may also therefore try to discover under which types of circumstances one sex may be more affected than the other.

Cortisol responders versus non-responders. A large number of the participants originally recruited into this study were subsequently classified as cortisol non-responders. Many of these non-responders were women. As already noted, previous studies that have analysed the results of cortisol responders and non-responders separately have found that there may be differences in memory performance between these groups (see, e.g., Buchanan & Tranel, 2008; Elzinga & Roelofs, 2005). Although such analyses were not run in this study because I was only interested in how elevated cortisol levels affected verbal declarative and working memory performances, future studies should make sure that such responder versus non-responder analyses comparisons are conducted in order to get a clearer overall picture of how exactly inter-individual responses to stress may affect cognitive performances.

In addition, in the current study it was clear that cortisol responders and non-responders could also be identified in the Relax groups (i.e., participants whose cortisol levels increased in response to the relaxation condition vs. those who showed stable or decreasing cortisol levels). Although previous studies do not appear to have identified such participants, it might prove interesting for future research to investigate how different cortisol responses in participants who undergo the control protocol affect cognitive functioning.

The type of declarative memory being studied. As already noted, previous studies have shown that the emotional valence of the to-be-remembered material can moderate the effects of

stress on declarative memory performance (Buchanan & Tranel, 2008; Domes et al., 2004; Tops et al., 2003). Specifically, previous literature shows that memory for positively valenced materials may be more negatively affected under conditions of stress than neutral material. In contrast, however, memory for negatively valenced material may be enhanced under conditions of stress. This result has been found specifically for recognition memory, where the ability to recognise positive material is decreased under stressful situations while the ability to recognise negative material is increased (Domes et al., 2004; Tops et al., 2003). As already mentioned, this differential ability to remember or recognise emotionally charged material compared to neutral material could have important consequences for every day functioning, because the stimuli encountered in real-life are often emotionally charged. Therefore, it would be more externally valid to investigate more thoroughly the effects of stress on memory for emotional material.

In addition to the moderating effects of the emotional valence of material on memory performance, it has been proposed that the type of encoding utilised by studies (e.g., intentional/explicit vs. incidental/implicit encoding) may be a moderating factor for the effects of stress on memory (Lupien et al., 1999). However, this proposal has not been thoroughly investigated and research appears to generally utilise the intentional encoding of to-be-recalled material (e.g., de Quervain et al., 2000, 2003; Domes et al., 2004; Kirschbaum et al., 1996, Study 1; Kuhlmann et al., 2005; Lupien et al., 1999; Wolf, Convit, et al., 2001; Wolf, Schommer, et al., 2001) more often than incidental encoding (e.g., Kirschbaum et al., 1996, Study 2; Lupien et al., 2002, Study 2). As a consequence, results are not definitive, and more research is needed in this area.

The stage of the declarative memory process being studied. Studies have also shown that the stage of the memory process at which the stressor is introduced can significantly moderate how the stressor affects cognitive performance (de Quervain et al., 2000). As noted previously, studies involving delayed free recall tasks have generally shown that this type of memory performance is not impacted by increased cortisol levels at the encoding phase of the memory process (de Quervain et al., 2000; Luethi et al., 2009; Nater et al., 2007; Wolf, Convit et al., 2001). These results are in contrast with research showing that increased cortisol levels at the retrieval phase of the memory process does appear to affect delayed free recall memory performance (de Quervain et al., 2000; Kuhlmann et al., 2005; Wolf, Convit, et al.,

2001).Although there are some contradictions to this pattern of results (e.g., Domes et al., 2004), they do appear to be fairly consistent across most studies.

The results in studies investigating the effect of stress on immediate free recall and delayed cued recall, however, appear to be far more mixed. Some research has found that these memory processes are affected by increased cortisol at the level of encoding (immediate free recall: Nater et al., 2007; Tops et al., 2003; delayed cued recall: Kirschbaum et al., 1996), whereas other research has found otherwise (immediate free recall: de Quervain et al., 2000; Wolf, Convit, et al., 2001; delayed cued recall: Lupien et al., 1999). In addition, results are equally inconsistent when cortisol levels are increased at the retrieval phase, with some studies finding impairments (delayed cued recall: de Quervain et al., 2003) and some not (delayed cued recall: Kuhlmann et al., 2005).

Thus, it is appears that the stage of the memory process at which the stressor is introduced can potentially moderate the effects of the stressor on the recall of the material. In addition, Lupien et al. (1999) state that it is important, if the study is being conducted on one day with cortisol levels being raised at encoding and then recall being tested soon afterwards, to ensure that the increased cortisol levels at encoding are not carrying over to the retrieval stage and thus actually affecting the memory process at this stage, thereby clouding conclusions. Although the current study supplies additional information that delayed cued recall is not affected by stress at the retrieval phase of the memory process, clearly more information is needed to fully elucidate how the stage of the memory process at which stress is experienced affects performance. This is an especially interesting question in terms of test and examination processes, where stress is generally felt at retrieval (i.e., during the writing of the actual exam), as it may help with understanding how people perform on such tasks.

Time of day. Although this study attempted to control for the circadian pattern of cortisol, clearly it does not reveal much about how time of day actually affects HPA axis functioning reactions to stress, and therefore what impact this variable may have on cognitive performance. It is apparent that more studies using different forms of memory (e.g., working memory) need to actively investigate what effects time of day has on stress and cognitive performance by running the same experimental protocol at a number of different times of day (such as was done in Maheu et al., 2005). Thus, the effects of time of day on stress on memory

functioning can be explicitly studied and it can be determined at which time of day stress has the most impact on cognitive performance. This could provide information about at which points in the day stress is most debilitating, and may therefore have important real-world implications.

Summary and Conclusions

The current study uncovered some interesting results, specifically in terms of the possible existence of sex differences in working memory performance under conditions of stress. Because sex differences in the effects of stress on memory performance have generally not been sufficiently investigated in previous research, this is a fairly novel finding which needs to be explored further in order to see if it is a finding that can be replicated under different circumstances with different variables being controlled for.

The current study has also identified a number of factors, including race, sex and time of day, which need to be taken into account in future research. Future studies should not only aim to better control for the many variables that may potentially impact results, but should also investigate these variables further. For instance, the current study showed sex differences in working memory performance between women in the late luteal phase of the menstrual cycle and men. However, whether these differences exist during other phases of the menstrual cycle is not addressed by the current design.

Unfortunately, because each study conducted in this area generally only controls for a subset of important variables, or uses their own (not necessarily standardized) tests to investigate memory performance, the ways in which all of the above discussed variables might interact together have not been sufficiently investigated. Thus, all of these factors need to be further researched in order to more fully understand the effect that a psychosocial stressor can have on different forms of memory performance.

It is clear that continued research into the effects of stress on memory performance is necessary. Results in this area have important consequences for further understanding human cognitive functioning and could potentially have important real-world impacts. People experience acute stressors in everyday life and they are often expected to function effectively under these (e.g., school and university examinations, giving academic or work-related presentations, and working in areas such as law enforcement, emergency medicine or fire fighting). Gaining a better understanding of how stress affects memory and other cognitive

domains and how these effects are moderated by individual or contextual differences, could help to better equip people in stressful situations to function more effectively.

References

Alderson, A. L., & Novack, T. A. (2002). Neurophysiological and clinical aspects of glucocorticoids and memory: A review. *Journal of Clinical and Experimental Neuropsychology, 24,* 335-355.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders (4th ed.)*. Washington, DC: Author.

Baddeley, A. D. (2001). Is working memory still working? *American Psychologist, 56,* 864-878.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50,* 7-15.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory manual (2nd ed.).* San Antonio, TX: Psychological Corporation.

Bischof, P. (2003, September). *The menstrual cycle*. Postgraduate Training Network for Research in Reproductive Health, The Geneva Foundation for Medical Education and Research. Retrieved May 6, 2008, from http://www.gfmer.ch/Endo/Course2003/Menstrual_cycle/Menstrual_cycle_Bischof.htm

Bleecker, M. L., Bolla-Wilson, K., Agnew, J., & Meyers, D. A. (1988). Age-related sex differences in verbal memory. *Journal of Clinical Psychology, 44,* 403-411.

Bonito Attwood, C. (2008). *The impact of acute psychological stress on spatial cognition.* Unpublished master's thesis, University of Cape Town, Cape Town, Western Cape, South Africa.

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

Buchanan, T. W., & Tranel, D. (2008). Stress and emotional memory retrieval: Effects of stress and cortisol response. *Neurobiology of Learning and Memory, 89,* 134-141.

Chong, R. Y., Uhart, M., McCaul, M. E., Johnson, E., & Wand, G. S. (2008). Whites have a more robust hypothalamic-pituitary-adrenal axis response to a psychosocial stressor than blacks. *Psychoneuroendocrinology, 33,* 246-254.

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (2007). Temporal dynamics of brain activation during a working memory task. *Nature, 386,* 604-607.

Cole, L. A., Ladner, D. G., & Byrn, F. W. (2009). The normal variabilities of the menstrual cycle. *Fertility and Sterility, 91*, 552-527.

de Kloet, E. R., Oitzl, M. S., Joëls, M. (1999). Stress and cognition: Are corticosteroids good or bad guys? *Trends in Neuroscience, 22*, 422-426.

de Quervain, D. J.-F., Henke, K., Aerni, A., Treyer, V., McGaugh, J. L., Berthold, T, et al. (2003). Glucocoricoid-induced impairment of declarative memory retrieval is associated with reduced blood flow in the medial temporal lobe. *European Journal of Neuroscience, 17,* 1296-1302.

de Quervain, D. J.-F., Roozendaal, B., Nitsch, R. M., McGaugh, J. L., & Hock, C. (2000). Acute cortisone administration impairs retrieval of long-term declarative memory in humans. *Nature Neuroscience, 3,* 313-314.

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin, 130,* 355-391.

Domes, G., Heinrichs, M., Rimmele, U., Reichwald, U., & Hautzinger, M. (2004). Acute stress impairs recognition for positive words – Association with stress-induced cortisol secretion. *Stress, 7,* 173-181.

Dozois, D. J. A., Dobson, K. S., & Ahnberg, J. L. (1998). A psychometric evaluation of the Beck Depression Inventory-II. *Psychological Assessment, 10,* 83-89.

Elzinga, B. M., & Roelofs, K. (2005). Cortisol-induced impairments of working memory require acute sympathetic activation. *Behavioral Neuroscience, 199,* 98-103.

Garde, A. H., & Hansen, Å, M. (2005). Long-term stability of salivary cortisol. *Scandinavian Journal of Clinical & Laboratory Investigation, 65,* 433-436.

Gillespie, B. L., & Eisler, R. M. (1992). Development of the feminine gender role stress scale [Abstract]. *Behavior Modification,* 16, 426-438. Abstract retrieved from http://bmo.sagepub.com/cgi/content/abstract/16/3/426

Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review, 12,* 229-246.

Henry, M. (2008). *The impact of acute psychosocial stress on false memory.* Unpublished honour's thesis, University of Cape Town, Cape Town, Western Cape, South Africa.

Jarrold, C., & Towse, J. N. (2006). Individual differences in working memory. *Neuroscience, 139,* 39-50.

Kelly, M.M., Tyrka, A. R., Anderson, G. M., Price, L. H., & Carpenter, L. L. (2008). Sex differences in emotional and physiological responses to the Trier Social Stress Test. *Journal of Behavior Therapy and Experimental Psychiatry, 39*, 87-98.

Kemeny, M. E. (2003). The psychobiology of stress. *Current Directions in Psychological Science, 14*, 124-129.

Kimura, D., & Hampson, E. (1994). Cognitive pattern in men and women is influenced by fluctuations in sex hormones. *Current Directions in Psychological Science, 3*, 57-61.

Kirschbaum, C., Kudielka, B. M., Gaab, J., Schommer, N. C., & Hellhammer, D. H. (1999). Impact of gender, menstrual cycle phase, and oral contraceptives on the activity of the hypothalamus-pituitary-adrenal axis. *Psychosomatic Medicine, 61*, 154-162.

Kirschbaum, C., Pirke, K-M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test' – A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology, 28*, 76-81.

Kirschbaum, C., Pirke, K-M., & Hellhammer, D. H. (1995). Preliminary evidence for reduced cortisol responsivity to psychological stress in women using oral contraceptive medication. *Psychoneuroendocrinology, 20*, 509-514.

Kirschbaum, C., Wolf, O. T., May, M., Wippich, W., & Hellhammer, D. H. (1996). Stress- and treatment-induced elevation of cortisol levels associated with impaired declarative memory in healthy adults. *Life Science, 58*, 1475-1483.

Kirschbaum, C., Wüst, S., & Hellhammer, D. (1992). Consistent sex differences in cortisol responses to psychological stress. *Psychosomatic Medicine, 54*, 648-657.

Kudielka, B. M., Hellhammer, D. H., & Wüst, S. (2009). Why do we respond so differently? Reviewing determinants of human salivary cortisol resposes to challenge. *Psychoneuroendocrinology, 34*, 2-18.

Kudielka, B. M., & Kirschbaum, C. (2005). Sex differences in HPA axis response to stress: A review. *Biological Psychology, 69*, 113-132.

Kudielka, B.M., Schmidt-Reinwald,A. K., Hellhammer, D. H., & Kirchbaum, C. (1999). Psychological and Endocrine Responses to Psychosocial Stress and dexamethasone/corticotrophin-releasing hormone in healty postmenopausal women and young controls: The impact of age and a two-week estradiol treatment. *Neuroendocrinology, 70*, 422-430.

Kudielka, B.M., Schmidt-Reinwald,A. K., Hellhammer, D. H., Schürmeyer, T., & Kirchbaum, C. (2000). Psychosocial stress and HPA functioning: No evidence for a reduced resilience in healthy elderly men. *Stress, 3*, 229-240

Kudielka, B. M., Schommer, N. C., Hellhammer, D. H., & Kirschbaum, C. (2004). Acute HPA axis responses, heart rate, and mood changes to psychosocial stress (TSST) in humans at different times of day. *Psychoneuroendocrinology, 29*, 983-992.

Kuhlmann, S., Piel, M., & Wolf, O. T. (2005). Impaired memory retrieval after psychosocial stress in healthy young men. *The Journal of Neuroscience, 25*, 2977-2982.

Luethi, M., Meier, B., & Sandi, C. (2009). Stress effects on working memory, explicit memory, implicit memory for neutral and emotional stimuli in healthy men. *Frontiers in Behavioral Neuroscience, 2*, 1-9.

Lupien, S. J., Gillin, C. J., & Hauger, R, L. (1999). Working memory is more sensitive than declarative memory to the acute effects of corticosteroids: A dose-response study in humans. *Behavioral Neuroscience, 113*, 420-430.

Lupien, S. J., Wilkinson, C. W., Brière, S., Ménard, C., Ng Yin Kin, N. M. K., & Nair, N. P. V. (2002). The modulatory effects of corticosteroids on cognition: Studies in young human populations. *Psychoneuroendocrinology, 27*, 401-416.

Lynn, R., & Irwing, P. (2008). Sex differences in mental arthimetic, digit span, and *g* defined as working memory capacity. *Intelligence, 36*, 226-235.

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences.* Stanford, CA: Stanford University Press.

Maheu, F. S., Collicut, P., Kornik, R., Moszkowski, R., & Lupien, S. J. (2005). The perfect time to be stressed: A differential modulation of human memory by stress applied in the morning or in the afternoon. *Progress in Neuro-Psychopharmacology & Biological Psychiatry, 29*, 1281-1288.

Maitland, S. B., Herlitz, A., Nyberg, L., Bäckman, L., & Nilsson, L.-G. (2004). Selective sex differences in declarative memory. *Memory & Cognition, 32*, 1160-1169.

Nater, U. M., Moor, C., Okere, U., Stallkamp, R., Martin, M., Ehlert, U., & Kliegel, M. (2007). Performance on a declarative memory task is better in high than in low cortisol responders to psychosocial stress. *Psychoneuroendocrinology, 32*, 758-763.

Oei, N. Y. L., Everaerd, W. T. A. M., Elzinga, B. M., Van Well, S., & Bermond, B. (2006). Psychosocial stress impairs working memory at high loads: An association with cortisol levels and memory retrieval. *Stress, 9*, 133-141.

Owen, A. M. (1997). The functional organization of working memory processes within human lateral frontal cortex: The contribution of functional neuroimaging. *European Journal of Neuroscience, 9*, 1329-1339.

Pillay, T. S., Haumann, C., Bonito Attwood, C., Omar, F., & Thomas, K. G. F. (2008). Salivary free cortisol reference range on the Roche E170 platform after collection with Salimetrics Sorbette® ("Eyespear"). *Annals of Clinical Biochemistry, 45*, 111.

Porcelli, A. J., Cruz, D., Wenberg, K., Patterson, M. D., Biswal, B. B., & Rympa, B. (2008). The effects of acute stress on human prefrontal working memory systems. *Physiology & Behavior, 95*, 282-289.

Psychological Corporation. (1998). *Wechsler Memory Scale Third Revision.* London: Harcourt Brace.

Sapolsky, R. (2004). *Why zebras don't get ulcers (3rd ed.).* New York: Owl Books.

Schoofs, D., Preuβ, D., & Wolf, O. T. (2008). Psychosocial stress induces working memory impairments in an *n*-back paradigm. *Psychoneuroendocrinology, 33*, 643-653.

Sherman, B. M., & Korenman, S. G. (1975). Hormonal characteristics of the human menstrual cycle throughout reproductive life. *The Journal of Clinical Investigation, 55*, 699-706.

Shimamura, A. P., Jurica, P. J., Mangels, J. A., Gershberg, F. B., & Knight, R. T. (1995). Susceptibility to memory interference effects following frontal lobe damage: Findings from tests of paired-associate learning. *Journal of Cognitive Neuroscience, 7*, 144-152.

Sliwinski, M. J., Smyth, J. M., Hofer, S. M., & Stawski, R. S. (2006). Intraindividual coupling of daily stress and cognition. *Psychology and Aging, 21*, 545-557.

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

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State–Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D., & Vagg, P. R. (1984). Psychometric properties of the STAI: A reply to Ramanaiah, Franzen, and Schill. *Journal of Personality Assessment, 48*, 95-97.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review, 99*, 195-213.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*, 652-654.

Stroud, L. R., Salovey, P., & Epel, E. S. (2002). Sex differences in stress responses: Social rejection versus achievement stress. *Society of Biological Psychiatry, 52*, 318-327.

Tops, M., van der Pompe, G., Baas, D., Mulder, L. J. M., den Boer, J. A., Meijman, T. F., & Korf, J. (2003). Acute cortisol effects on immediate free recall and recognition of nouns depend on stimulus valence. *Psychophysiology, 40*, 167-173.

Uhart, M., Chong, R. Y., Oswald, L., Lin, P-I., & Wand, G. S. (2006). Gender differences in hypothalamic-pituitary-adrenal (HPA) axis reactivity. *Psychoneuroendocrinology, 31*, 642-652.

Uttl, B., Graf, P., & Richter, L. K. (2002). Verbal paired associates tests limits on validity and reliability. *Archives of Clinical Neuropsychology, 17*, 567-581.

van den Bos, R., Harteveld, M., & Stoop, H. (2009). Stress and decision-making in human: Performance is related to cortisol reactivity, albeit differently in men and women. *Psychoneuroendocrinology, 34*, 1449-1458.

Vedhara, K., Hyde, J., Gilchrest, I. D., Tytherleigh, M., & Plummer, S. (2000). Acute stress, memory, attention and cortisol. *Psychoneuroendocrinology, 25*, 535-549.

Wechsler, D. (1987). *Wechsler Memory Scale-Revised.* San Antonio, TX: Psychological Corporation.

Weiss, E. M., Kemmler, G., Deisenhammer, E. A., Fleischhacker, W. W., & Delazer, M. (2003). Sex differences in cognitive functions. *Personality and Individual Differences, 35,* 863-875.

Wolf, O.T. (2003). HPA axis and memory. *Best Practice & Research Clinical Endocrinology and Metabolism, 17,* 287-299.

Wolf, O. T., Convit, A., McHugh, P. F., Kandil, E., Thorn, E. L., De Santi, S., et al. (2001). Cortisol differentially affects memory in young and elderly men. *Behavioral Neuroscience, 155,* 1002-1011.

Wolf, O. T., Schommer, N. C., Hellhammer, D. H., McEwen, B. S., & Kirschbaum, C. (2001). The relationship between stress induced cortisol levels and memory differs between men and women. *Psychoneuroendocrinology, 26,* 711-720.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18,* 459-482.

Appendix A

Declarative Memory Cued Recall Task: VPA Word Lists

Day 1: Trial 1.

Study Order:	***Recall Order:***
Frog-neck	Rose (flower)
Metal-iron	Fruit (apple)
Foot-tree	Lion (circus)
School-grocery	Room (face)
Fruit-apple	Coal (year)
Colour-blue	Metal (iron)
Hill-ring	School (grocery)
Baby-cries	Hill (ring)
Obey-inch	Frog (neck)
Crush-dark	Cabbage (pen)
Lion-circus	Prison (thief)
Girl-sign	Colour (blue)
Coal-year	Bank (milk)
Room-face	Girl (sign)
Prison-thief	Obey (inch)
Rose-flower	Foot (tree)
Cabbage-pen	Baby (cries)
Bank-milk	Crush (dark)

Day 1: Trial 2.

Study Order:	***Recall Order:***
Bank-milk	Obey (inch)
Frog-neck	Bank (milk)
Room-face	Hill (ring)
School-grocery	Prison (thief)
Coal-year	Crush (dark)
Girl-sign	Coal (year)
Rose-flower	Lion (circus)
Prison-thief	Room (face)
Colour-blue	Colour (blue)
Obey-inch	Foot (tree)
Baby-cries	Girl (sign)
Fruit-apple	Baby (cries)
Metal-iron	Metal (iron)
Cabbage-pen	Frog (neck)
Foot-tree	Fruit (apple)
Crush-dark	Rose (flower)
Hill-ring	School (grocery)
Lion-circus	Cabbage (pen)

Day 2.

Recall Order:

Bank-milk

Frog-neck

Room-face

Colour-blue

School-grocery

Coal-year

Girl-sign

Prison-thief

Rose-flower

Obey-inch

Baby-cries

Fruit-apple

Metal-iron

Cabbage-pen

Lion-circus

Foot-tree

Crush-dark

Hill-ring

Appendix B

Declarative Memory Recognition Task: VPA Word List

FROG – NECK	**SCHOOL – GROCERY**
Frog – Pond	School – Children
Throat – Neck	Shop – Grocery
Frog – House	School – Cream
Wall – Neck	Hair – Grocery
METAL – IRON	**FRUIT – APPLE**
Metal – Steel	Fruit – Vegetable
Ore – Iron	Pear – Apple
Metal – Pillow	Fruit – Holiday
Shoe – Iron	Ship – Apple
FOOT – TREE	**BANK – MILK**
Foot – Shoe	Bank – Money
Leaf – Tree	Cow – Milk
Foot – Window	Bank – Tablet
Chair – Tree	Sofa – Milk
CABBAGE – PEN	**HILL – RING**
Cabbage – Boiled	Hill – Climb
Ink – Pen	Jewel – Ring
Cabbage – Mat	Hill – Scissors
Bed – Pen	Car – Ring

OBEY – INCH

Obey – Command

Ruler – Inch

Obey – Green

Cinema – Inch

COAL – YEAR

Coal – Fire

Month – Year

Coal – Book

Bonnet – Year

GIRL – SIGN

Girl – Boy

Traffic – Sign

Girl – Plastic

Razor – Sign

ROSE – FLOWER

Rose – Thorn

Vase – Flower

Rose – Alarm

Club – Flower

ROOM – FACE

Room – House

Eyes – Face

Room – Dart

Basket – Face

PRISON – THIEF

Prison – Arrest

Steal – Thief

Prison – Bee

Water – Thief

BABY – CRIES

Baby – Dummy

Tears – Cries

Baby – Light

Muscle – Cries

LION – CIRCUS

Lion – Roar

Monkey – Circus

Lion – Glue

Key – Circus

CRUSH – DARK

Crush – Ice

Night – Dark

Crush – Key

Hammer – Dark

COLOUR – BLUE

Colour – Paint

Sky – Blue

Colour – File

Nail – Blue

Appendix C

Consent Form

Informed Consent to Participate in Research and Authorization for Collection, Use, and Disclosure of Protected Health Information

This form provides you with information about the study and seeks your authorization for the collection, use and disclosure of your protected health information necessary for the study. The Principal Investigator (the person in charge of this research) or a representative of the Principal Investigator will also describe this study to you and answer all of your questions. Your participation is entirely voluntary. Before you decide whether or not to take part, read the information below and ask questions about anything you do not understand. By participating in this study you will not be penalized or lose any benefits to which you would otherwise be entitled.

1. Name of Participant ("Study Subject")

_____

2. Title of Research Study

The impact of acute psychological stress on cognitive functioning

3. Principal Investigators, Ethics Committee, and Telephone Numbers

Kevin G. F. Thomas, Ph.D.
Department of Psychology
University of Cape Town
021-650-4608

Robyn Human, B.Soc.Sc. (Hons)
Masters Candidate
Department of Psychology
University of Cape Town
021-788-5536

Michelle Henry, B.Sc., B.Soc.Sc. (Hons)
Masters Candidate
Department of Psychology
University of Cape Town
021-551-6534

Faculty of Health Sciences
Research Ethics Committee
Room E52-24 Groote Schuur Hospital Old Main Building
Observatory 7925
Tel: 021-406-6338
Fax: 021-406-6411
Email: lamees.emjedi@uct.ac.za

4. What is the purpose of this research study?

The purpose of this research study is to better understand how exposure to acute psychological stress affects cognitive functioning. More specifically, we are interested in individual differences in cognitive responses to acute psychological stress.

5. What will be done if you take part in this research study?

This study requires you to take part in two research sessions on two consecutive days. On the first day you will be required to complete a number of memory-based tasks. On the second day you may be required to complete a 20-minute presentation which will be followed by another series of memory-based tasks. Throughout the second day of the study your levels of stress will be assessed through the collection of self-report data, heart rate measurements, skin conductance measurements and saliva samples with the aid of a cotton swab. These saliva samples will be used to analyse levels of cortisol, a stress hormone.

6. What are the possible discomforts and risks?

If you are one of the participants selected to complete the 20-minute presentation, you may be placed in a mildly stressful situation involving public speaking. There are no other discomforts and risks associated with participation in the study.

7. What are the possible benefits of this study?

One major benefit of this study is that scientists and society in general, will have better understanding of the effects of acute psychological stress on cognitive functioning. This knowledge can then be applied to many different individuals and situations, including students who are taking exams, business managers who have to present to their boards, and so on.

8. Can you withdraw from this research study and if you withdraw, can information about you still be used and/or collected?

You may withdraw your consent and stop participation in this study at any time. Information already collected may be used.

9. Once personal information is collected, how will it be kept confidential in order to protect your privacy and what protected health information about you may be collected, used and shared with others?

Information collected will be stored in locked filing cabinets or in computers with security passwords. Only certain people - the researchers for this study and certain University of Cape Town officials - have the legal right to review these research records. Your research records will not be released without your permission unless required by law or a court order.

If you agree to be in this research study, it is possible that some of the information collected might be copied into a "limited data set" to be used for other research purposes. If so, the limited data set may only include information that does not directly identify you.

10. Signatures

As a representative of this study, I have explained to the participant the purpose, the procedures, the possible benefits, and the risks of this research study; the alternatives to being in the study; and how the participant's protected health information will be collected, used, and shared with others:

_____ _____
Signature of Person Obtaining Consent and Authorization Date

You have been informed about this study's purpose, procedures, and risks; how your protected health information will be collected, used and shared with others. You have received a copy of this form. You have been given the opportunity to ask questions before you sign, and you have been told that you can ask other questions at any time.

You voluntarily agree to participate in this study. You hereby authorize the collection, use and sharing of your protected health information. By signing this form, you are not waiving any of your legal rights.

_____ _____
Signature of Person Consenting and Authorizing Date

Please indicate below if you would like to be notified of future research projects conducted by our research group:

_____ (initial) Yes, I would like to be added to your research participation pool and be notified of research projects in which I might participate in the future.

Method of contact:
Phone number: _____
E-mail address: _____
Mailing address: _____
_____
_____

Appendix D

Comparison of Cortisol Levels Before and After Participant Exclusion

The following figures show the differences in cortisol changes from baseline to post-experimental manipulation between the total sample (excluding the participant in the SF group with the very high cortisol levels) and the final sample. It is hoped that these figure give an indication of the effect of removing the cortisol non-responders from the final statistical analyses.

Figure D1 shows that while in both samples the Stress groups showed an increase in cortisol levels after the experimental manipulation, this increase is obviously greater in the final sample (total sample: M =3.06 ± 3.82; final sample: M = 4.91 ± 3.53). In the Relax groups, the final sample shows the obviously expected decrease in cortisol levels after the experimental manipulation (M = -0.46 ± 0.85), while the total sample shows a small increase (M = 0.13 ± 1.50).

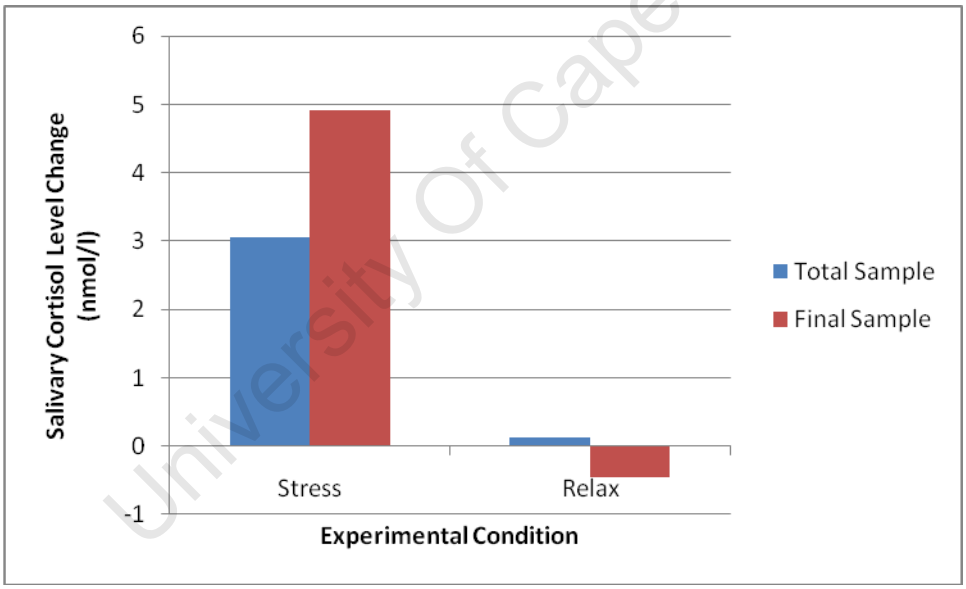

Figure D1. Comparison of change in cortisol levels from baseline to post-experimental manipulation for total and final sample Stress and Relax groups.

Figure D2 shows that when the four groups are looked at separately, it is clear that participants in the final sample SF group (M = 3.65 ± 2.72) showed a much greater cortisol increase than participants in the total sample SF group (M = 1.68 ± 2.85). Participants in the final sample SM group (M = 6.01 ± 3.86) also show a greater increase than participants in the total sample SM group (M = 4.98 ± 4.23), although this difference is smaller than that between the two SF groups. In addition, it can be seen that both the participants in the total

sample RF (*M* = 0.20 ± 0.90) and RM (*M* = 0.05 ± 1.99) groups show an increase in cortisol levels while participants in the final sample RF (*M* = -0.22 ±0.38) and RM (*M* = -0.68 ± 1.09) groups both show a decrease in cortisol levels.

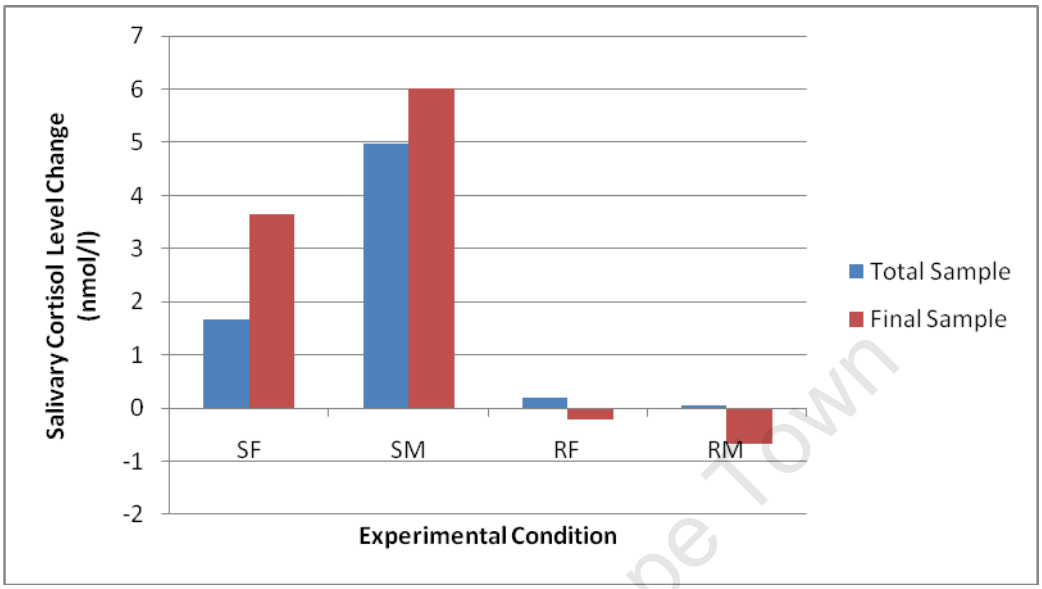

Figure D2. Comparison of change in cortisol levels from baseline to post-experimental manipulation for total and final sample SF, SM, RF and RM groups.

Appendix E

Accuracy of Phase of Menstrual Cycle Estimation for Total Female Sample

Of the 21 female participants who completed the study but were not included in the final sample, 15 were tested in the desired phase of their menstrual cycle. Of the remaining six, one had already begun her period before the second session, but had not informed us of this fact. The other five were tested more than 6 days before the start of their menstrual cycles.

Overall, the participants were fairly accurate in their predictions of when the menses phase of their cycle would begin. Of the total number of female participants who completed the study (n = 48), more than half of them (n = 31) were either correct or only one day off with their predictions.

Figure E1 shows how many days away from the start of their periods the participants were on Day 2 (negative numbers indicate days prior to period). Figure E2 shows the accuracy of the female participants in predicting when the start of their next period would be (negative numbers indicate periods beginning before the predicted date).

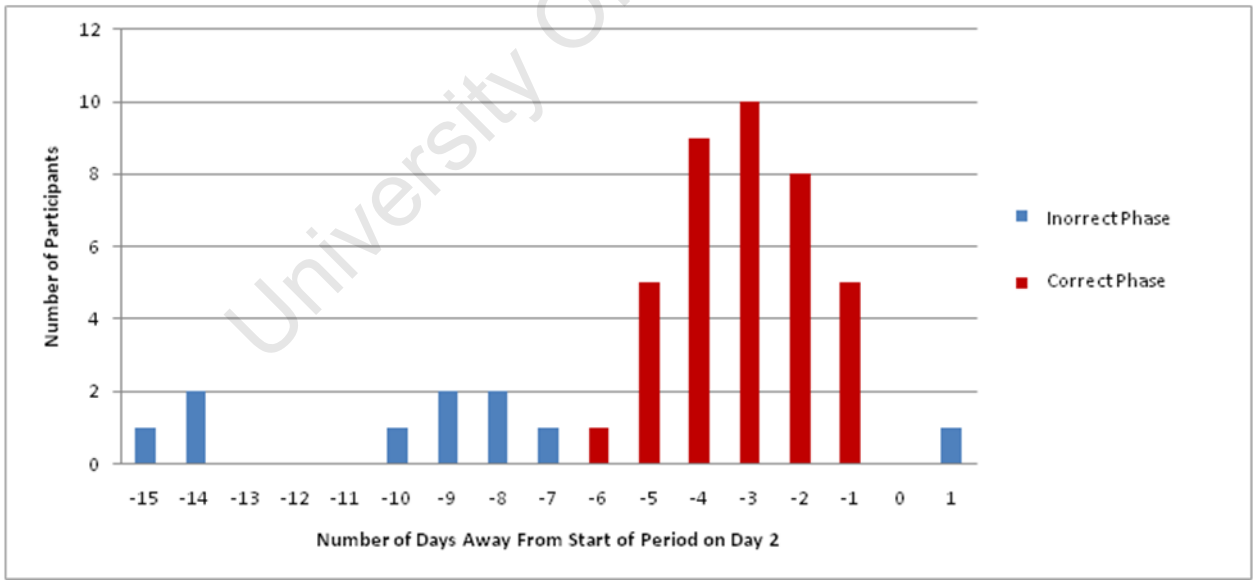

Figure E1. Number of total female participants in the correct and incorrect phases of their menstrual cycle on Day 2.

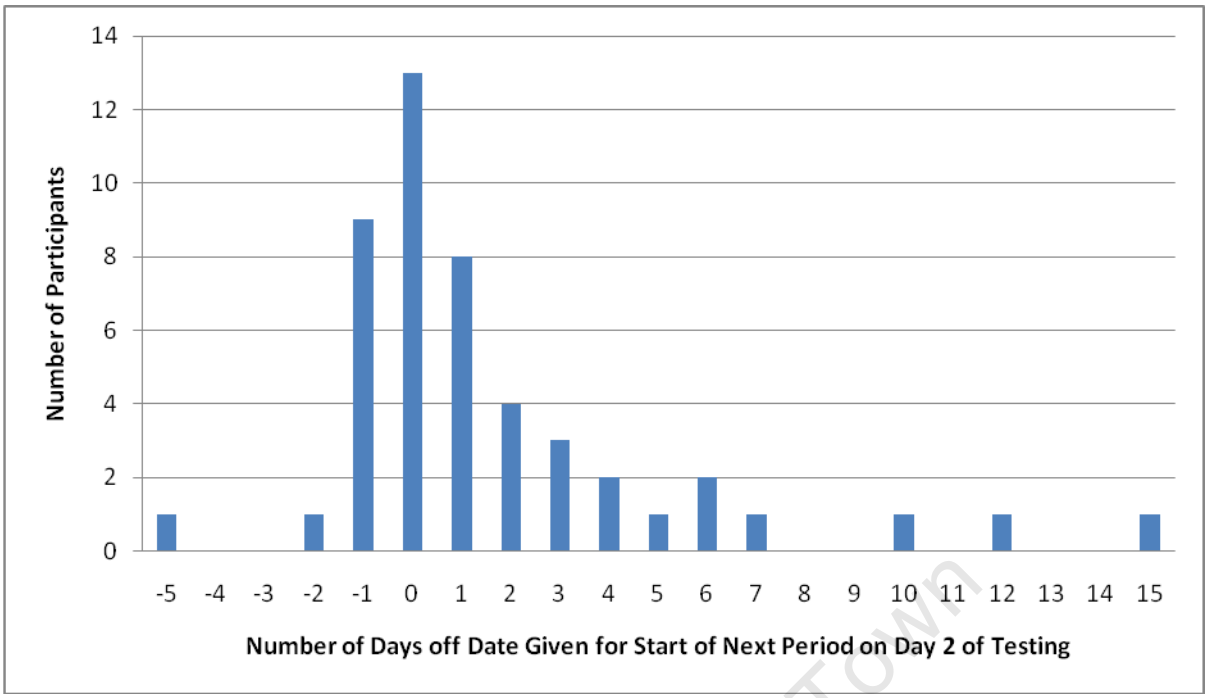

Figure E2. Accuracy of female participants in predicting one what date their next period would begin.