Emotional working memory training, work demands, stress and anxiety in cognitive performance and decision-making under uncertainty

Author: Amanda J. Heath
Supervisor: Laszlo Harmat
Examiner: Gary N. Burns
Acknowledgements

Thanks to Dr. Susanne Schweizer (University College London), for access to and support with the $n$-back and FM training, and Stroop and digit span tests, and without which it would have been very difficult to repeat the methods used in her studies. Thanks also to my supervisor, Dr. Laszlo Harmat and the Master programme leader Dr. Andrejs Ozolins (LNU), for patience and support; Prof. Gijsbert Stoet (Leeds Becket University), regarding advice on PsyToolkit; Tom Pollard, Dr Jason Rutter, and Dr. Faye Skelton (Napier University), for technical help on MATLAB; Dr Magda Osman (Queen Mary University London), for advice on variants of the Iowa Gambling Task; Annika Herrmann and Meike Petermann, for helpful and thorough feedback on the first manuscript; Prof. Gary Burns for specific feedback, and to Steven Milosevic for last-minute proofreading. Thanks most of all to all of the participants who gave around six hours of their time in total, over the course of 16 days, to complete the entire study, half of whom have been to ‘hell $n$-back!’ I salute you!
Abstract

The study seeks to bring together literature on decision-making, the effects of work-related demands and stress, and individual differences in trait anxiety on near and far transfer effects of emotional working memory training (eWM). A sample of 31 students and working participants underwent emotional working memory training through an adaptive dual n-back method or a placebo face match training task for 14 days. Pre- and post-training measures were taken of a near transfer task, digit span, medium transfer measure of executive control, emotional Stroop, and a far transfer task of decision-making under uncertainty, the Iowa Gambling Task (IGT). In line with previous studies, eWM was expected to show gains in transfer task performance between pre- and post-training, and, especially for those scoring high on trait anxiety and workplace measures of stress demands (taken from COPSOQ), for whom there is more scope for improvement in emotional regulation. Gains in emotional Stroop specifically were further expected to show support for the effects of eWM training on emotional well-being in addition to decision-making. Results fell short of replicating previous work on transfer gains, though interference effects in Stroop did lessen in the eWM training group. Relationships between work demands, anxiety, stress and performance in the training itself, reinforce previous research showing that work stress and anxiety lead to cognitive failures, highlighting the importance of intervention studies in the organizational field, but they were not linked to benefits of the training. Resource and methodological limitations of the current study are considered, especially those involved in conducting pre-post designs and cognitive testing online.
Demanding jobs that cause stress and strain in working life can not only lead to burnout and be detrimental to health and well-being generally, but potentially deleterious to performance and decision-making in work tasks. Judgement and decision-making (JDM) are important features of many roles in organizations, especially leadership positions (Ganster, 2005; Hambrick, Finkelstein, & Mooney, 2005). In jobs where lapses of attention and decision-making can have potentially very serious consequences (for example, emergency room workers, air traffic controllers, police), these failures are of particular concern, and interventions that target the cognitive and neural architecture underlying JDM, and seeking to improve stress regulation and executive control, should be of great interest to organizational psychologists. The present study tested a cognitive intervention to improve decision-making, mental control and emotion regulation in students and professionals, and relationships to anxiety, work demands and stress, with implications for judgement and performance in organizational contexts.

Background

Job demands and stress. Stress has variedly been defined in terms of two broad aspects, or as a combination of the two: 1) the features of the external environment that act on an individual, and/or 2) the individual’s psychological, physiological, and behavioural responses to external demands, threats and challenges (Ganster & Rosen, 2013). External events trigger internal reactions to ultimately affect well-being and judgement. These events are referred to as stressors and the individual responses to them as strains. One can differentiate between environmental stressors like those in physically demanding jobs, and psychologically demanding or psychosocial ones (Ganster & Rosen, 2013), and the current study is concerned with the latter. I adopt Ganster & Rosen’s definition of work stress as the “process by which workplace psychological experiences and demands (stressors) produce both short-term (strains) and long-term changes in mental and physical health” (p. 1088).

Job stress as a response to challenges or demands, is conceptualised as a set of transactional processes with cognitive appraisal of external demands and coping resources determining individual reactions (Ganster, 2005; Lazarus & Folkman, 1984; Lazarus, 1993). High demands interact with appraisal to produce arousal that acts on cognition and emotion, and which can, in turn, impair judgement and decision-making. High job demands can also lead to poorer performance simply through increasing task demand (Ganster, 2005), as harder tasks are, by definition, more difficult to complete successfully. The most popular theories of workplace stress identify certain aspects of the working environment as key to ill-health and poor outcomes like low job satisfaction, sickness absence, burnout and high turnover.
The most widely influential model of work stress is Karasek’s (1979) demand-control, latterly, demand-control-support model (Karasek, 1989). It predicts that control over stressors and the buffer of social support mitigate the effects of job demands. Another popular model, along similar lines, is the effort-reward-imbalance model (Siegrist, 1996), and describes how internal motivations and external pressures act together, with perceived rewards buffering their influence on job satisfaction, health outcomes and well-being. Both models have received empirical support in the literature, and are the most notable examples of several with overlapping assumptions (Van der Doef & Maes, 1999).

Psychosocial job demands come in different forms, cognitive (attentional and mental), quantitative (load), and emotional (as in emotional labour). Models of workplace stress would predict that jobs with high demands will require greater resources to cope adequately. Some of these resources are internal (coping styles, self-efficacy), while others are external (supportive colleagues, resources to be able to do one’s job adequately). Of most interest in the current study are those job demands that place stress on executive function, the higher cognitive processes that enable planning, forethought, and goal-directed action (Shields, Sazma, & Yonelinas, 2016), and which encompass executive control and emotion regulation processes. In jobs that require high precision, constant attention, and in which performance errors can lead to serious consequences, the study of interventions to improve coping responses can make a valuable contribution to the organizational field. While previous research has been conducted that aims to increase intrapersonal resources, such as self-efficacy (e.g., Hahn, Binnewies, Sonnentag & Mojza, 2011), emotional intelligence (e.g., Nikolaou & Tsaousis, 2002), or train effective coping strategies (see Lamontagne, Keegel, Louie, Ostry & Landsbergis, 2013), interventions that target improvements in cognitive resources are not common. Perhaps exceptions being cognitive-behavioural therapy that addresses reasoning processes (van der Klink, Blonk, Schene & van Dijk, 2001), and implementation intentions (Morgan & Harris, 2015) that attempt to turn unrealised intentions into concrete form. Both of these approaches operate on cognitions at the explicit level, and not at the implicit level. The current study is an attempt to fill the gap, and address the control side of the demand-control-support triangle using an approach drawn from the recent literature on working memory training and mental control. As such, it is to my knowledge, the first attempt to apply a cognitive neuroscience intervention that aims to enhance implicit executive functioning, to impact performance and emotion regulation linked to job demands and stress.
Very demanding jobs that induce chronic stress may lead to exhaustion and burnout with decrements in executive functions, attention and memory (Deligkaris, Panagopoulou, Montgomery, & Masoura, 2014; Gilboa, Shirom, Fried, & Cooper, 2005; Van Der Linden, Keijser, Eling & Van Schaijk, 2007). However, they can also pose a level of challenge that is beneficial. Then et al. (2014) studied data from a German sample of 2,725 participants, linking performance on tests of cognitive function with mental work demands, classified using the O*NET standardised job descriptions of cognitive abilities at work. High job demands were associated with better cognitive function, independent of education and intelligence. The authors concluded that jobs with high cognitive demands extend capacity and have a protective effect on cognition in the long term. They link their results to findings on the reduced risk of dementia in later life for those in more intellectually challenging careers. However, whether job demands or stressors are perceived as a challenge or a threat will depend upon many individual factors (Lazarus & Folkman, 1984). This is where we should expect to see the strongly interactive effects of individual differences affecting the interplay of demands, control and support.

The university is an organisation, and a demanding programme of studies can mimic the demands seen in the workplace. Theoretical models of stress should also apply to the student situation. In one such empirical example, Sieverding et al. (2013) extended Karasek’s demand-control model to identify high demands and low decision latitude as predictors of student stress and low student satisfaction. Similarly, Mokgele & Rothmann (2014) applied the job demands-resource model (Demerouti, Bakker, Nachreiner, & Schaufeli., 2001) to the student situation, and concluded that high study demands promote burnout, with the concomitant deleterious effects on health and well-being, and resources playing a buffering effect. Wöfel et al. (2016) also found that both depression and anxiety in a student population was related to structural conditions predicted by the job demands-resource model. Subjective ratings of study demands and available resources were positively correlated with anxiety and depression, with study demands, time latitude and social support contributing 16% of the variance in anxiety scores. The prevalence of anxiety and depression in their German sample was higher than that in the general population, showing that study demands can lead to significant stress despite a rather different structure of working patterns from most paid work.

The present study also examined support for the theoretical models of work stress to a student group, comparing students and working participants.

**Decision-making under stress.** There has been a plethora of research in the area of judgement and decision-making generally since influential work of Kahneman and Tversky
in the 1970s began to map the personal and contextual factors that impact everyday reasoning. Much research has focused specifically on decision-making under conditions of *uncertainty* - where individuals lack knowledge of consequences of a specific choice, and under those of *ambiguity* in which outcomes cannot be fully specified and or the probabilities are unknown or uncertain. *Risk* refers to situations where the probability of a possible outcome is known, at least in part, and many studies have examined estimation of probabilities, finding common biases and errors in judgement (Tversky & Kahneman, 1974; Kahneman, 2011; Morgado, Sousa, & Cerquiera, 2015), and sensitivities of context, such as those highlighted through the framing literature (Kahneman & Tversky, 1982). One of the main characteristics observed of decision-making under stress is in processing of risk. Parallel neural circuits that link the cerebral cortex and basal ganglia are believed to underlie decision-making itself, and stress affects these systems through glucocorticoid receptors (Sousa & Almeida, 2012). Alterations to the dopamine firing system, which affect prefrontal cortex functioning, impair executive control. Consequently, stress leads to less inhibited and more reward seeking responses, biases in the estimation of probabilities of loss/gain especially when large rewards are on offer, and delay discounting or favouring smaller but immediate rewards over larger, long-term rewards (Pabst, Brand & Wolf, 2013; Starcke & Brand, 2016). Phenomena such as hypervigilance for threatening stimuli, and attention narrowing, impulsively selecting an alternative without consideration of the full range of alternatives (Gok & Atsan, 2016), interact with biases and heuristics like availability, seizing on what easily comes to mind (Finucane, Alhakami, Slovic, & Johnson, 2000). These combine to orchestrate riskier decisions under stress, ‘fast’ system 1 thinking, less analytical consideration of viable alternatives, and faulty cost-benefit analysis and probability estimation (Gok & Atsan, 2016; Starcke & Brand, 2016). Though these heuristics and biases might be adaptive in some situations (de Kloet et al, 2008; Roozenendaal, 2002), and indeed physiological responses to stress in the short-term prepare the body for action, it remains that when making complex and consequential decisions in the modern world, and perhaps especially in the workplace, these reactions can be detrimental to information processing and judgement (Janis & Mann, 1977). The many negative health implications of prolonged activation of stress hormones are also well-documented (see Sousa & Almeida, 2012, for a review of the research) with consequences for the individual and organizations beyond mere short-term performance deficits.

The relationship between cognition and emotion is two-fold in relation to organizational concerns. Firstly, cognitive processes, specifically executive function (mental
control, emotion regulation, attention and memory) mediate the experience of stress and the individual response to it, as described in the transactional approach (Lazarus, 1991; Lazarus & Folkman, 1984). Secondly, once stress and strain is experienced through the former transaction, how the individual deals with the resulting arousal is further determined by their executive functioning abilities, and, in particular, the ability to regulate emotion. This will be at the heart of the current research intervention.

Anxiety, through placing extra task irrelevant demands upon capacity, impairs working memory function in particular, which is a key component of executive function. A recent meta-analysis by Moran (2016) concluded impairments of both self-reported and induced laboratory stress on working memory on various tasks and across modalities, and another meta-analysis by Shields, Sazma, & Yonelinas (2016) drew similar conclusions, finding evidence of impairment specifically in cognitive flexibility, though the evidence suggested more nuanced effects on inhibition. The effects of chronic long term stress, through prolonged activation of the hypothalamus-pituitary-adrenal (HPA) axis and release of stress hormones, is revealed in structural brain changes to the areas underlying executive function in humans and animals (Dias-Ferreira, et al., 2009; Lupien, McEwan, Gunnar, & Heim, 2009). Cognitive interventions that aim to improve executive function would appear a fruitful avenue to improve decision-making in stressful conditions, especially if they target emotion/fear regulation systems in the brain, and potentially augment cognitive control of stress.

Stress and job performance. Several studies have shown the adverse effects of acute and chronic stress on cognitive function in occupational groups specifically, corroborating other literature on stress and decision-making. Effects observed include judgement errors, poorer impulse control, and general decision-making decrements from reliance on biases and heuristics (Morgan, Doran, Steffian, Hazlett, & Southwick, 2006; Regehr, LeBlanc, Jelley, & Barath, 2008; Taverniers, Ruysseveldt, Smeets, & Grumbkow, 2010). As Gutshall et al. (2017) point out, long-term exposure to on-the-job stress can lead to altered perceptions, impact problem-solving, interfere with recall, and affect the interpretation of the intentions and responses made by others. All of these have ramifications for job performance, satisfaction and organizational effectiveness. In professions that involve regular demands under acute stress, such as police and law enforcement, poor decisions in the heat of the moment can have fatal consequences. Decrement to decision-making and judgement can be expected especially when making decisions under uncertainty, where outcomes are unknown or only partially knowable, where prior knowledge and experience may not help, and where
analytical thinking skills come to precedence. According to the principle of bounded rationality, which fits well with demand-control-support models, decision makers are always constrained in their judgements by the tractability of the problem, the cognitive, time and other resources available, and decisions will be determined by these factors rather than deployment of optimal, rational-analytical methods (Simon, 1956). It follows that in situations which involve clear target threats, or in cases where one can fall back on learned heuristics based on experience, stress-induced biases are functional, and associated with performance enhancement.

Indeed, empirical evidence with occupation groups suggests just this. For example, Akola & Mendes (2012) found a positive relationship between cortisol levels and performance in a ‘shoot-don’t shoot’ task using police officers, where heightened vigilance for threat was facilitative. Demands which are challenging, such as making decisions under time pressure, can also be facilitative. Payne, Bettman, & Johnson (1988) found that time pressure resulted in functional narrowing of attention on central, important aspects of the situation, and reliance on prior knowledge and heuristics worked well in given contexts. The authors also observed contextual variation in the strategies deployed, not simply repetitive use of any one heuristic or bias. However, executive function, and specifically working memory processes, are required especially in decision-making tasks requiring extensive processing of decision options and their consequences, when decision makers cannot fall back on learned associations and prior experience, or discern clear perceptual patterns or strategies (Del Missier, Mäntylä, & De Bruin, 2012). But because biases like hypervigilance and attention narrowing cause biased encoding and limited consideration of alternatives (Hammond, 2000), they are, as mentioned, especially problematic when making judgements under uncertainty, and which is common in many high stress professions. These twin processes also interfere with the encoding and recall of information, and several studies have shown detriments to short-term and long-term memory under stress in occupational groups such as police officers (Gutshll et al., 2017; Lewinski, Dysterheft, Priem, & Pettitt, 2016; McCarty & Lawrence, 2016). This fits with a body of literature suggesting impairments of stress in laboratory decision-making and memory tasks (Moran, 2016). Therefore, the current study specifically examines decision-making under uncertainty where these cognitive processes are key to risky decisions. Further, the choice of task is determined by the need for a measure that relies on emotion-related control processes relevant to decision-making under pressure.
Decision-making and individual differences. The Iowa Gambling Task (IGT) is a well-established laboratory task that is argued to call upon executive control and emotion-related regulation processes (Buelow & Suhr, 2009). It is a measure of risky decision-making in which participants try to maximise winnings by making selections between four decks of cards, two of which yield high gains but also high losses and result in an overall loss if favoured, and two of which yield low gains but low losses and result in an overall gain. The IGT simulates both uncertainty of premises and outcomes – the number of trials (100) is not revealed to participants in advance, and trials on which losses occur are determined randomly and not predictable. Despite being devised primarily for use with clinical populations, the IGT has been found to demonstrate a great deal of individual variation in performance in normal populations (Bechara & Damasio, 2005; Dunn, Dalgleish, & Lawrence, 2006). This does imply systematic individual differences that account for variation in attentional, memory and interpretive biases. There are also gender effects in the IGT, with men shown to favour more the low-reward, advantageous decks more than women (van den Bos, Homberg, & de Visser, 2013). Finucane, Alhakami, Slovic, & Johnson (2000) propose that risk evaluations in decision-making are made primarily based on affective feedback – whether we have positive or negative associations with an alternative, something they refer to as the ‘affect heuristic.’ Drawing on somatic marker theory, Bechara and colleagues argue that performance on IGT is involves covert, pre-cognitive emotional signals that adaptively guide decision-making even before explicit knowledge about the task is available to consciousness (Bechara, Damasio, Damasio, & Anderson, 1994) – these somatic ‘warning’ signals should lead to avoidance of the bad decks in normal subjects, even when respondents have no explicit insight into the premises and probabilities associated with the task. However, evidence with variants of the task suggest some explicit processes are implicated in IGT performance (Dunn et al., 2006; Buelow & Surh, 2009), and, over the course of trials, participants can gain awareness of which are the good and bad decks, hold the relative gains and losses in mind, make an estimation of the costs/benefits and so on. These are skills calling upon attention and working memory, and so IGT performance seems to be determined by both ‘hot’ (emotional) and ‘cold’ (analytical) reasoning (Buelow & Surh, 2009).

Anxiety is defined as negative emotional arousal in response to perceived threat (Lazarus, 1991), and trait anxiety (TA) is the personality disposition towards an anxious state in the anticipation of threat. Unsurprisingly this has been found to affect decision-making on tasks such as the IGT (Miu, Heilman, & Houser, 2008). High TA individuals might be expected to show special sensitivity to somatic markers (Damasio, Tranel, & Damasio,
1991), especially given other findings linking high TA to pre-attentional cognitive biases (Mathews & Macintosh, 1998). Indeed, high TA is associated with risk aversion and pessimistic risk appraisals (Maner et al., 2007; Mitte, 2007). High TA individuals show impaired IGT performance and demonstrate increased somatic markers (measured through skin conductance) i.e. anticipatory responses, for advantageous trials in the IGT. This implies defective modulation of somatic feedback that should normally lead to avoidance of the bad decks (Miu et al., 2008). TA is assumed to be a stable disposition (Endler & Kocovski, 2001; Gray & McNaughton, 2003), with neural substrates in the prefrontal cortex and amygdala revealed by functional and structural brain differences (Miu et al., 2008). Arguably enhanced attention to threat and subsequent negative somatic markers impair general processing of risks on the disadvantageous decks, with more attention given to the larger rewards, despite larger losses and overall disadvantage (delay discounting). A simpler explanation might be that the high trait anxiety leads to higher cognitive load and greater burden on executive function, and thus general interference with decision-making neural-circuits. Either way, the link between trait anxiety and faulty executive function processing is quite well established.

As Schweizer, Grahn, Hampshire, Mobbs & Dalgleish (2013) point out, humans vary enormously in their ability to keep a cool head in emotionally charged situations. The capacity to remain goal-focused when selecting plans under relevant affective distraction is a marker of success in many domains of life. Deficits in affective cognitive control are characteristic of many psychiatric disorders and have highlighted the role of the mediating neural circuits (Morgado et al., 2015; Schweizer et al., 2013). Pharmacological and physical interventions have been used with these clinical populations. However, the neural circuits and brain regions underlying affective cognitive control also underlie executive processes more generally and, in particular, working memory (Sari, Koster, Pourtois, & Derakshan, 2016; Schweizer, Hampshire & Dalgleish, 2011; Schweizer et al., 2013; Pabst et al., 2013), and so it follows that training interventions that focus on improving executive and working memory functions more generally might have a knock on effect for emotional regulation, not to mention their direct involvement in decision-making processes.

*Can executive control and working memory be trained?* There has been much interest in the past decade in the potential benefits of working memory training on general cognitive performance in other untrained, i.e. transfer, tasks (Buschkuehl, Jaeggi, & Jonides, 2012; Jaeggi, Buschkuehl, Jonides & Perrig, 2008; Karbach & Verhaeghen, 2014). Working memory is a central component of cognition and highly correlated, though distinct from, general intelligence (Lawlor-Savage & Goghari, 2016). Particular attention has been given to
dual-task n-back training, which has shown small but significant gains in fluid intelligence (Au, Sheehan, Tsai, Duncan, Buschkuehl, & Jaeggi, 2015), performance on tasks of executive control (Sari, Koster, Pourtois, & Derakshan, 2016) and standard measures of attention (Kundu, Sutterer, Emrich, & Postle, 2013). The findings are not without contention, however, and some authors have argued support only for benefits on near-transfer tasks (of working memory capacity) rather than far-transfer tasks (like fluid intelligence tests) (Melby-Lervåg, Redick, & Hulme, 2016). It is also unknown if any of the effects found from training are sustainable in the long-term, while Redick (2015) points out that the gains found in some studies are down to decreases in post-training performance in the control group rather than increases in the experimental group. The Jaeggi and Redick groups have published meta-analyses with conflicting conclusions (Au et al., 2015; Redick & Lindsey, 2013). Certainly methodological limitations further obscure conclusions, with some studies using passive and some active control groups and the latter finding weaker effects, and frequently small samples under 20 in each group, with biases towards student samples at elite educational institutions. There also appears to be differences in effectiveness of training depending upon the offer of financial incentives to complete (Au et al., 2015), and throw light perhaps on the importance of motivation to engage with the training task. However, some compelling evidence in the body of published work does exist of gains from dual-task n-back, including that from interventions with elderly groups and those with mild cognitive impairments (Salminen, Frensch, Strobach, & Schubert, 2015; Xin, Lai, Li, & Maes, 2014). The idea that training executive processes could have far reaching benefits is an appealing one, and seems theoretically well-founded, if perhaps difficult to establish due to the onerous nature of training studies requiring high effort by participants outside of the lab, and associated with high drop out and plagued by general impracticalities.

The n-back task. The n-back task can be presented as a single or dual task in two simultaneous modalities. In the dual version, it involves monitoring visual or visuospatial and auditory stimuli simultaneously, and deciding if the current trial matches the stimulus presented n trials back, with the n varying dependent on performance. This later factor appears to be crucial, i.e. that the training is highly challenging and adaptive, with the level of n-back always dynamically determined by current performance level of the participant. This keeps the participant perpetually engaged at their cognitive limit (Au et al., 2015). Plasticity is argued to be driven by a mismatch between demands and cognitive capacity (Lövdén, Bäckman, Lindenberger, Schaefer, & Schmiedek, 2010). Increasing demand on executive
function pushes the limits of one’s abilities, rather like working out at a gym, it expands capacity and improves cognitive functioning.

The effects of WM training, especially in medium and far transfer tasks not directly related to WM, is further supported by brain imaging work which shows changes in task-related effective brain connectivity, specifically in frontoparietal and parieto-occipital networks that are thought to underlie performance in a range of tasks (Kundu et al., 2013). Indeed, Román et al. (2017) used structural brain imaging after training to uncover evidence of enhanced connectivity in regions supporting working memory, interference resolution, inhibition, and task engagement. In a modification of the standard dual-n-back procedure on which the current study is based, Schweizer, Hampshire & Dalgleish (2011) and Schweizer, Grahn, Hampshire, Mobbs & Dalgleish (2013) presented emotional stimuli – emotional faces in a 4x4 grid, simultaneous to auditory presented threat words (rape, dead, hate, etc.). They found improvement in a working memory task after 20 days of eWM training in comparison to non-affective WM training and placebo feature match training, with transfer benefits in visuospatial fluid intelligence and emotion regulation (Schweizer et al., 2011). Schweizer et al.’s (2013) fMRI data revealed processing gains through changes in activation in regions underlying the frontoparietal demand network. The authors concluded that training using the emotional version of n-back in particular stimulates the prefrontal cortical connections to the limbic system, which downregulate emotional distress (Schweizer et al., 2013). Thereby improving performance especially in tasks requiring avoidance of distraction or interference of emotion-laden stimuli. It follows, that challenging and repetitive training, employing the affective cognitive control system which underpins emotion regulation, should further augment performance under stress and anxiety through the same downregulatory mechanisms. To reinforce the idea that executive processes implicated in emotion can be trained, Sari et al (2016) recently used a form of n-back working memory training, over a three-week period, to reduce anxiety and improve WM performance in a group of high trait anxiety individuals, i.e. those for whom faulty emotion regulation impairs decision-making under stress. These data suggest the potential of emotional WM training to improve real world decision-making under conditions of emotional stress and strain in occupations which involve high demands on executive function.

**Current study**

The present study used Schweizer et al.’s emotional variant of n-back training for 14 days, alongside an active control group in a pre-test – post-test design that attempts to improve decision-making under uncertainty, and demonstrate particular benefits to
individuals with high work demands and stress. Schweizer and colleagues’ method had a limitation in the control group, where participants performed an object feature matching task, introducing a potential confound that the gains in the experimental group were due to simple exposure to faces, not through improvements in emotion regulation. The current research corrects for this limitation by using an active control group from Schweizer’s more recent unpublished work, that presents the same face stimuli as the n-back task, while making minimal demands on working memory and executive function.

The effects of training were assessed on a near transfer task of working memory capacity, digit span, a medium transfer executive function test, emotional Stroop, and a far transfer task of decision-making under uncertainty, the Iowa Gambling Task, and linked performance and gains to scores on state and trait anxiety, work demands and stress. It was expected that the training intervention would work to improve decision-making though enhancing executive functioning, and have particular benefits with trait anxious individuals and those suffering from high workplace demands and stress. There should be more scope for improvement in transfer tasks for those who currently have impaired executive function through high anxiety or high stress. Support for this would not only confirm theoretical ideas about the underlying cognitive and neural mechanisms of decision-making under stress, and therefore add to our understanding, but also, if the training is successful especially with trait anxious individuals and those suffering from high workplace demands and stress, it suggests a cost-effective intervention that can be used in the workplace, and may be of value especially in acutely demanding professions such as the police, air traffic control, and emergency responders where optimal workplace judgement is crucial and where effective emotion regulation is required on a daily basis. Those who score high in work demands that in particular place stress on attentional, decision-making and memory processes, should see greater benefits of the training, as arguably this should augment everyday performance in the job. A comparison of professionals and students was made, while measuring perceived study load/challenge/strain in the same manner as work demands, and which, on the basis of previous research, was expected to show the same link between demands and gains in training as for those in full-time employment (Wöfel et al., 2016).

Research aims. Firstly, I aimed to replicate transfer gains in previous studies using emotional n-back training and shed light on the theory of why these gains occur by examining performance on a near, medium and far transfer task. Specifically, using a measure of decision-making known to be sensitive to stress and anxiety, with the expected influence of TA, work stress and demand further supporting notions that eWM training
enhances executive function and emotion regulation in particular. Secondly, link any observed gains in emotional $n$-back training to work demands, and to observe improvements in cognitive stress symptoms and work-related anxiety.

Decision-making under uncertainty in a well-established task, IGT, that measures risk taking (which is known to increase in those under stress), is expected to improve after eWM training. It is further expected that emotional Stroop response latencies will decrease after eWM training, offering a potential causal mechanism for training gains in decision-making and any interactional effects of demanding work. A further question is whether these gains are higher for those with trait anxiety (TA) and high reported workplace stress, given studies have shown an impairment of TA on tasks reliant on working memory and emotional regulation. Digit span performance, as closely related to working memory, should improve after working memory training if the training has any effect at all.

The interaction of trait anxiety should support the idea that failure of emotion regulation has a detrimental effect on cognition generally and decision-making specifically, especially tasks requiring inhibition of interference or irrelevant information. The training intervention, if successful especially with trait anxious individuals, not only confirms theoretical ideas about the underlying cognitive and neural mechanisms of decision-making under stress, but also suggests the training can help some individuals in particular to perform better and regulate responses under stress. Positive gains related to cognitive stress and workplace demands would then give particular impetus to the fruitfulness of further research that directly examines workplace performance as a result of cognitive training interventions, and especially in professions of high cognitive demand.

**Hypotheses**

H1: based on previous literature on working memory training, eWM training will show gains in performance on a near transfer task of working memory, digit span.

H2: based on previous studies using the emotional variant of $n$-back, eWM training will show gains in performance on a medium transfer ‘gold-standard’ measure of executive control, emotional Stroop, showing the effects of the training on emotion regulation and control processes.

H3: based on the premise that working memory training using affective stimuli acts on the emotional circuits connecting frontoparietal and limbic system regions involved in emotional control and regulation as found in Schweizer et al. (2013), eWM training will show gains in performance on a far transfer task of decision-making under uncertainty, the
Iowa Gambling Task, established to require emotion regulatory processing of somatic markers, and influenced in particular by stress and anxiety.

H4: Trait anxious individuals will be expected to show greater gains from emotional working memory training in all tasks, with this acting as a moderator, partly due to a ceiling effect in degree of improvement with the low TA individuals, and partly down to learning adaptive emotional regulation strategies to remedy their existing weaker emotion regulation.

H5: High work demands will be associated with greater gains in transfer task performance, as the training should augment everyday job performance where high cognitive and attentional demands feature heavily.

H6: Transfer performance will show a negative relationship to trait anxiety (TA) overall, and high TA participants will record higher response latencies to emotional Stroop. reflecting delayed processing of emotion-laden information, and lower gains from making riskier choices in IGT in particular (Miu et al, 2008).

H7: For the same reasons, cognitive stress and SA will correlate positively with high response latencies to emotional words in Stroop and negatively IGT scores.

H8: High TA scores are expected to correlate with higher reports of cognitive stress, due to the detrimental effects of anxiety on executive function.

H9: State anxiety and cognitive stress should decrease after eWM training in comparison to the placebo training group, as participants are better able to regulate their emotional responses to stress.

**Method**

**Participants**

*Sampling.* Some authors have noted that findings in the WM training and JDM literature might be explained by differences in sample demographics (Moran, 2016), with many samples consisting of high achievement students drawn from elite educational institutions. It was thought beneficial to try to replicate the findings of previous WM training studies with a more heterogeneous sample than in previous studies. Advertising was aimed at those who are currently in full time, part time or self-employed work via targeted social media groups. The implications of success in improving decision-making and emotion regulation in people currently in the workplace, rather than just those in intensive education, can allow us see if wider real world benefits are possible from the training. A student sample
was also recruited for comparison. The tests were in English, and therefore English-speaking individuals were recruited, though not necessarily native speakers.

Completion and dropouts. In total 187 volunteers signed up using the online form to register an interest in the study, and were emailed pre-training logins and instructions. Fifty-nine of these completed the pre-training, and, of these, 35 went through the training programme and completed post-training tests. A valid programme of training consisted of 11 of 14 sessions, and 4 participants were excluded from the post-data analyses after falling short of this, leaving 31 participants, 15 in the control group and 16 in the experimental group.

The mean age of participants completing pre-training was 34.1 (SD = 12.3), range 18-67, and of those completing a valid training programme was 38.1 (SD = 12), which was in the control group 39 (SD = 12.42) [n = 15] and in the experimental group 37.4 (SD = 11.8) [n = 16], range 23 -59. Thirty-nine women and 20 men completed the pre-training, and 23 females and 8 men completed a valid training programme.

Occupation data showed that participants were mainly from white collar backgrounds, teaching, administration, and healthcare were pre-dominant job areas, with students making up just under half of the initial pre-training sample. Most participants were either in full time employment [n =21] or full time studies [n = 23], and the final figures completing valid training were 13 and 10 respectively, with others unemployed, in part-time employment or on sabbatical. Over 74% of the sample was educated to at least bachelor level. English was the first language for 28 participants and 31 spoke some other language as a first language, and this was 17 and 15 respectively in the valid training completion group.

Participants were volunteers recruited through posters around the university campus, and advertisements placed on social media and in web forums, further spread through word of mouth sharing on Facebook by friends and acquaintances. In addition to university social media groups, ads were placed on Facebook groups for English-speaking expats in Sweden, and as a general call on Twitter with hashtags #psychology, #cognition, #brain training. Ads were also placed on a national student forum in the UK (studentroom.com), and through posting the sign-up form on SurveyTandem.com, a site dedicated to researchers exchanging participation in survey based research. This site is used by students, researchers and academics to obtain survey responses. A financial inducement of being entered into a lottery draw for a monetary prize of 700 SEK/£60 first prize and 300 SEK/£25 second prize was offered to those completed a valid program of training, which was a minimum of 11 training sessions out of the 14, the 80% used in previous studies. The number of sessions that
constituted a valid programme was not revealed to participants in advance, and instead they were told if they missed a day they could carry on the next day, but if they missed too many sessions they would not be eligible for the prize draw.

**Design**

A randomised controlled design was chosen, with half the participants allocated to a placebo training, and half to the treatment/experimental group. It had a mixed factor, pre-test-post-test design. Trait and state anxiety were measured by the State-Trait Anxiety Inventory, with SA measured pre- and post-training. Work demands and cognitive stress were measured by subscales of the Copenhagen Psychosocial Stress Questionnaire (COPSOQ). Near transfer effects were measured by a digit span test of working memory capacity. An emotional Stroop task was deployed as a gold-standard measure of executive functioning (Schweizer et al., 2011), with response latencies measured in milliseconds. The effects of training on decision-making under uncertainty were measured by performance on the Iowa Gambling Task in hypothetical currency. All materials were in English, though Swedish adjectives were provided as additional information for STAI.

**Materials**

*Training.* The *emotional WM training* group used the dual n-back task of Schweizer et al (2013), presenting auditory and visuospatial threat related words and faces with negative emotional expressions in a 4x4 grid. The n targets varied from 1 upwards. In addition to general threat related words (‘hate’, ‘fear’), some words related to the concept of stress and worry in particular. Examples of the threat-related words include ‘distressed’, ‘afraid’, ‘overwhelmed’, ‘burdened’ and which are relevant to the concepts of stress and anxiety. *Placebo training* involved a facial feature matching task that makes minimal demands on working memory and taken from Schweizer’s recent as yet unpublished research (Schweizer, personal communication). This requires participants to match two panels/grids of faces containing different individuals and different facial expressions, and these two grids appeared side by side on the computer screen, so involve visual matching. Performance on the face match task was adapted to performance, and went up to 12 simultaneous faces to match. Participants completed the training online, and it was hosted on the Cambridge Brain Sciences (CBS) platform.

*State-Trait Anxiety.* Trait and State anxiety was measured by the State Trait Anxiety Inventory (STAI, version Y1/2) consisting of 40 items (20 for each) rated on a 4-point scale of ‘almost never’ to ‘almost always.’ Examples of items include “I worry too much over something that really doesn’t matter” and “I take disappointments so keenly that I can’t put
them out of my mind.” The STAI is a widely used measure and found to have high reliability and validity on various measures (Spielberger et al, 1983). Swedish translations of the adjectives were included in brackets after the English question, to help Swedish speakers with any of the more obscure words, and taken from an established translation of the STAI (as used in Forsberg & Björvell, 1993). At the end of both the state and trait subscales, participants were asked an additional question, to rate the extent to which their anxiety level is influenced by work status or studies on a 5-point scale.

COPSOQ. Workplace demands and cognitive stress were measured by subscales of the COPSOQ, which were chosen to reflect workplace demands on cognitive and attentional resources relevant to the training, and additionally the four items measuring symptoms of cognitive stress. The Copenhagen Psychosocial Stress Questionnaire (COPSOQ) was first developed in the late 1990s by the Danish National Research Centre for the Working Environment and developed over a decade (Kristensen, Bjørner, Christensen, Borg, 2004; Kristensen, Borg, & Hannerz, 2002; Kristensen, Hannerz, Høgh, & Bor, 2005; Kristensen, Pejtersen, Kristensen, Borg, & Bjørner, 2010). The full version aims to cover the whole range of psychosocial work environment stressors, resources and coping factors. It draws inspiration from common features of several theoretical models of workplace stress, such as Karasek’s demand-control-support model and Siegrist’s effort-reward-imbalance model (Nübling et al, 2006), while not being tied to one. The original full version contains 141 items on 5-point scales (Kristensen & Borg, 2003), while a revised version was published in 2010 with several items omitted, including those on sensorial demands at work, and new subscales included (Pejtersen, Kristensen, Borg, & Bjørner, 2010). The questionnaire has been translated into numerous languages and shows good validity and reliability across a range of occupational samples, and been adopted as a standard measure of psychosocial work environment in several countries (Bjørner & Petersen, 2010; Pejtersen et al., 2010; Thorsen & Bjørner, 2010).

Items from the subscales on job demands and cognitive stress were chosen from COPSOQ (see Appendix A). The demand subscales measured in the current study were quantitative demands (work load and time demands), cognitive demands (concentration, decisions, memory, coming up with ideas), sensorial demands (attention, concentration, precision), and demands for responsibility (financial, well-being, and physical consequences for others). It should be noted that the items on sensory demands were taken from the original COPSOQ version 1, though these have been removed from the second version of the questionnaire published in 2010 because they could either be inferred from the job title or
refer to unchangeable aspects of the job for some professions (Pejtersen et al., 2010). These items were included as they relate very closely to attentional demands in the work place, and which are of high relevance to skills being trained and the study hypotheses. The job demands characteristics measured by COPSOQ partially overlap with items from other well-used scales measuring job demand such as the job content questionnaire (Karasek et al., 1998). In addition to job demands, the four items on cognitive stress symptoms were used, these ask participants to rate symptoms of cognitive stress over the past 4 weeks on 5-point scales (How much of the time during the past 4 weeks have you... had problems concentrating, had difficulty with remembering, had difficulty in taking decisions, found it difficult to think clearly). High demand scores have been related to long-term sickness absence (Pejtersen et al., 2010; Rugulies, Aust & Pejtersen, 2010), and they were chosen for particular relevance to the current study as they assess demands related to attention, memory and executive function generally. These items also are well suited to a mixed sample of professionals and students, and the items can equally apply to studies as well as work, without adjustment in wording.

**Digit Span and Stroop.** Forward digit span was tested through serial presentation of a list of single digit numbers on the screen, which participants had to remember and were required to type back correctly in order of presentation. Increasingly more digits are presented each trial until errors are made at which the number of items constitutes the individual’s span. Forward digit span is an accepted measure of verbal short-term working memory, and closely related to fluid intelligence (Kane, Hambrick, & Conway, 2005). Both this and the emotional Stroop task was that used in Schweizer et al. (2011), and involved the presentation of words of three categories of sad, angry, or happy. The words were superimposed upon faces with neutral or emotional facial expressions. Sometimes the facial expressions of the photos were congruent with the valence of the presented word (the word “irate” and an angry face), and sometimes incongruent (“sorrowful” presented with a happy face), which provided affective interference that should slow responses. The participants were instructed to press the arrow keys to indicate sad, angry, happy and correctly categorise the words as quickly as possible. Response latencies for the different types of trials was recorded. Emotional Stroop is considered a gold-standard measure of executive functioning, emotion regulation and avoidance of emotional distraction (Schweizer et al., 2011). Both tasks were hosted on the CBS platform.
Figure 1. A schematic diagram of the Iowa Gambling Task. Participants select cars from four decks, and aim to maximise gains from a loan of $2000. Turning each card carries an immediate reward ($100 in decks A and B and $50 in decks C and D) but sometimes also losses (from Bechara et al., 2005, ©Trends in Cognitive Sciences).

**IGT.** Decision-making under uncertainty was measured before and after training by performance on the Iowa Gambling Task (IGT) (Bechara, Damasio, Damasio, & Anderson, 1994). It involves a computerized card dealing task where the aim is to maximise monetary gains. Participants click on one of four decks to draw a card which either results in a gain or both a gain and a loss. Two of the decks contain high reward trials ($100) but also high reward losses ($250) and which result in an overall loss in favoured, while two of the decks yield lower rewards ($50) but lower losses ($50) and result in an overall gain if favoured (see Figure 3). Participants were instructed to maximise hypothetical winnings by selecting decks that provide the most gains, starting at 2000 currency, which they are told is a loan that must be paid back at the end. Feedback during the trials normally leads individuals to begin to favour the low reward decks which are more advantageous, but reward seeking and risky decision-making favours the high reward decks. There are 100 trials, though participants are not informed of this. Total loss-gain on the task was measured in hypothetical dollars. The task was presented using PsyToolkit (Stoet, 2017).

**Demographic characteristics.** Age, gender, first language, occupation/study subject, work status (full or part-time, student, other), and educational background, were elicited through items hosted on Google Forms, which also hosted the sign up form, briefing and debriefing materials, and which linked the different hosting platforms used to present materials/tests.
Procedure

Interested participants followed a URL to a briefing and sign-up form. They were then emailed unique study logins, to ensure anonymity, and pre-training instructions and URL. In the briefing participants were informed that the study was on ‘brain training’ to improve mental control, and that previous research had found benefits of brain training for IQ, stress regulation, memory and attention. They were further informed about the financial incentives for completing a valid training program to encourage participation, and told that the training was based on working memory and used emotional stimuli that trained memory and attention.

Volunteers were tested using computer-based software online in their own time. During pre- and post-training participants were asked to complete the tests in a fixed order, which was the emotional Stroop and digit span, followed by demographic questions including employment status (at time 1 only), the state anxiety inventory, the trait anxiety inventory (time 1 only), COPSOQ items on work demands (time 1 only) and cognitive stress, and finally the IGT. During post-training participants were also asked to give subjective evaluations of the effects of the training they had undertaken, and the final question in the study asked them to guess which training group they had been in.

Training. On completion of pre-training, participants were emailed a specific training login and URL to the CBS site, and instructed to perform training every day for 14 days, starting the day after the pre-training session. The experimental group did the emotional working memory (eWM) dual n-back training task, which is a dynamic dual modality n-back task, and requires participants to remember the stimuli n positions back whilst monitoring both modalities during each trial. Figure 2 represents a block in the training in both visual and auditory conditions where n = 1 back. A picture of the face can appear in one of 16 positions in the 4x4 grid during each trial. Simultaneously to the visual presentation participants hear auditory words and are required to indicated with a key press if the trial was a target trial or not. Targets were visual or auditory and presented in blocks of 20+n trials (picture-word pairs) with 6 target trials per modality each block. In the visual example, target C is the visual target because it appears in the same grid position as the previous trial (n = 1 back in this example) - participants are asked to ignore the face of the actor in the picture and focus on spatial position, and in the auditory stream the target is D as ‘evil’ was present 1 back. The n-back position is determined by performance, always starting with n = 1, if participants get three or more consecutive trails correct then it moves to 2, 3, and so on, with no maximum. Similarly, errors lead to n-back being decreased on the next block. The task lasted 20-30
minutes depending upon performance, though a ‘Quit’ button appeared after 10 minutes which allowed participants to cut short training. Participants were encouraged to do the full training to get the most benefits, and in practice most participants completing a valid number of training sessions did so by completed a full training for each or almost all days.

Figure 2. The emotional dual-task n-back task, sample training block of n-back=1. Stimuli with a highlighted border represent target stimuli for the current block. In this example, there is a match because, for the visuospatial modality, the current face appears in the same location as the face 1-position back; and for the auditory target, the word (“EVIL”) is the same as the word one-back (from Schweizer et al., 2013, ©The Journal of Neuroscience).

In the placebo training group, participants completed a facial feature matching task, presented with separate panels of faces in a grid and asked to press a key to indicate matching or non-matching panels with a mouse click. The number of panels increases and decreases with performance level but is limited to maximum 12.
Figure 3. Face feature match task used for the control placebo training in the current study. The presented trial is a mismatch due to a different face in the middle of the second row. Number of faces was from 1 to 12 depending on performance.

In both groups, email reminders of training were sent out each day, with a copy of the login information, URL, and reminder of the current day of training. At the end of 14 days, participants were emailed the login and links for the post-training, digit span, emotional Stroop, state anxiety, and COPSOQ items on cognitive stress. There were also asked a series of subjective evaluation questions about the effects of training rated on 7-point scales (see Appendix B). Participants then completed the Iowa Gambling Task, after which, a final study question asked them to guess what training group they had been in and indicate one of three responses: ‘experimental/working memory’, ‘placebo/control’ or ‘not sure’. Participants were then linked to the debriefing which thanked them for participation and explained the different training conditions and the study purpose in more detail. It also provided further information on the training and support information on stress and anxiety. Participants who had been in the ‘placebo’ training were emailed logins for the n-back training, so they could gain any benefits of the treatment group.

Ethical approval was obtained through the Departmental procedure for Master work requiring self-assessment and approved by supervisor. The participants were all adults and able to provide informed consent which they did through a checkbox on the pre-training form. Participants were informed of confidentiality and right to withdraw, that only summary and not individual data will be used in written publications of the data and names kept in a separate file with code numbers, and only retained to communicate regarding the prize draw
for valid training completion. Participants were briefed and debriefed fully, though blind to which training group they were in until completion. No physical risks were considered to result from the training, and, as the study was done online, participants completed it in their chosen environment. Participants were asked to complete questions on state and trait anxiety, which may have brought to mind an existing condition. Participants were forewarned of being asked questions on anxiety, and at the end provided with links to support groups and online resources for anxiety and work stress. Participants were informed in advance that training would involve emotional stimuli. At debriefing, participants allocated to the placebo training group were assured of the importance of their data to the research by a personal email, and given a login and access to the n-back training software so that they were able to experience any benefits of the training and not be disadvantaged in the long-term if the training is found to be beneficial.

Results

Data analyses and data screening. All data analyses were performed using SPSS version 25. Initial data screening involved checking for potential entry errors with unusually large values, and the replacement of missing values with a coding value, with SPSS instructed to use listwise deletion. These were mainly from missing post-training data, though a few missing data points occurred for pre-train Stroop and digit span, and so these cases are excluded from the relevant analyses. Data were screened for outliers, and in IGT post-train scores one case in the experimental group was removed from post training analyses, as it was 3 standard deviations from the mean, and examination of the trials showed the participant had selected deck A for the first 95 out of the 100 trials, moving to deck B for the final 5 and never selecting deck C or D. In the post-training Stroop test, one case in the control group had unusually long response times, over 2.6 seconds on average, and trials that were over 6 seconds long, especially when compared to their pre-train Stroop RTs which were very close to the average for the sample. This data point was replaced by the mean in the post-training Stroop analyses. Data met acceptable benchmarks for normality (Field, 2009). Assumptions such as homogeneity of variances in ANOVA and t tests, and of equality of covariance matrices were met for all tests unless otherwise specified.

Completion of valid training. Independent samples t tests revealed that those completing valid training programmes had lower pre-training state anxiety t (46.18 corrected) = 2.17, P = .04, than non-completions. Completions also had lower TA scores than those who did not t (47.33 corrected) = 2.15, P = .04, There was a tendency for non-completions to have
higher cognitive stress (low scores reflect high stress) measured at pre-training ($P = .06$). Descriptive statistics are in Table 1.

Table 1

*Descriptive statistics for those completing and not completing a valid training programme on state anxiety, trait anxiety, and cognitive stress*

<table>
<thead>
<tr>
<th>Variables</th>
<th>Training</th>
<th>$n$</th>
<th>$M$</th>
<th>$SD$</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA pre</td>
<td>Complete</td>
<td>32</td>
<td>33.06</td>
<td>9.65</td>
<td>29.58, 36.54</td>
</tr>
<tr>
<td></td>
<td>Incomplete</td>
<td>27</td>
<td>39.59</td>
<td>11.63</td>
<td>34.27, 44.91</td>
</tr>
<tr>
<td>TA</td>
<td>Complete</td>
<td>32</td>
<td>37.25</td>
<td>11.22</td>
<td>33.74, 41.83</td>
</tr>
<tr>
<td></td>
<td>Incomplete</td>
<td>27</td>
<td>45.15</td>
<td>15.07</td>
<td>39.19, 51.11</td>
</tr>
<tr>
<td>Cognitive stress pre</td>
<td>Complete</td>
<td>31</td>
<td>13.32</td>
<td>3.40</td>
<td>12.08, 14.57</td>
</tr>
<tr>
<td></td>
<td>Incomplete</td>
<td>27</td>
<td>11.44</td>
<td>4.12</td>
<td>9.81, 13.08</td>
</tr>
</tbody>
</table>

*Pre-test data and scale reliabilities.* Independent samples $t$ tests showed there were no significant pre-training differences between the experimental and control training groups on SA, TA, or the transfer tasks. Thus the training groups did not differ to begin with.

Reliabilities for the STAI were very high, for trait anxiety $\alpha = .96$ [$N = 59$], and for the state anxiety items pre-training $\alpha = .91$ [$N = 59$], and SA post-training $\alpha = .95$ [$n = 36$]. COPSOQ alphas were also high and are reported further on with their analyses in Table 7.

*Intercorrelations between study variables.* Descriptive statistics for the main study variables are presented in Table 2, and intercorrelations between main study variables in Table 3. There were no signs of interrelationships between performance scores on the different transfer tasks. Several relationships were noted with performance on the training, however. Both the maximum and average level of $n$-back achieved during training was positively associated with performance on digit span (pre- and post-training) and IGT performance (pre-training), and the correlations at post-training were in a similar direction but not significant. This suggests a positive association between performance on span and on IGT and $n$-back on the whole.
### Table 2

**Descriptive statistics for transfer tasks, trait-state anxiety, and training performance**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Digit span pre</th>
<th>Digit span post</th>
<th>Stroop pre</th>
<th>Stroop post</th>
<th>IGT pre</th>
<th>IGT post</th>
<th>SA pre</th>
<th>SA post</th>
<th>TA</th>
<th>Max. n-back</th>
<th>Av. n-back</th>
<th>FM score</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>6.84</td>
<td>7</td>
<td>1134</td>
<td>1043</td>
<td>2173</td>
<td>2323</td>
<td>33.06</td>
<td>35.41</td>
<td>37.78</td>
<td>6.06</td>
<td>4.58</td>
<td>1250</td>
</tr>
<tr>
<td>$SD$</td>
<td>1.61</td>
<td>1.69</td>
<td>316</td>
<td>319</td>
<td>913</td>
<td>1049</td>
<td>9.65</td>
<td>11.63</td>
<td>11.21</td>
<td>1.52</td>
<td>1.22</td>
<td>382</td>
</tr>
<tr>
<td>$n$</td>
<td>32</td>
<td>31</td>
<td>29</td>
<td>30</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>16</td>
</tr>
</tbody>
</table>

### Table 3

**Correlations between transfer task performance pre- and post-training, state and trait anxiety scores, and training performance**

<table>
<thead>
<tr>
<th>Variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Digit Span Pre</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Digit Span Post</td>
<td>.54**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Stroop Incongruent RT Pre</td>
<td>-.28</td>
<td>-.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Stroop Incongruent RT Post</td>
<td>-.10</td>
<td>-.15</td>
<td>.48*</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. IGT Pre</td>
<td>.25</td>
<td>.21</td>
<td>-.09</td>
<td>-.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. IGT Post</td>
<td>.14</td>
<td>.78</td>
<td>-.08</td>
<td>-.25</td>
<td>.59**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. SA Pre</td>
<td>-.19</td>
<td>-.13</td>
<td>.47*</td>
<td>.02</td>
<td>-.11</td>
<td>-.02</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. SA Post</td>
<td>-.05</td>
<td>-.06</td>
<td>.09</td>
<td>-.07</td>
<td>.11</td>
<td>-.07</td>
<td>.49**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. TA</td>
<td>-.17</td>
<td>-.17</td>
<td>.19</td>
<td>.06</td>
<td>-.35*</td>
<td>.38*</td>
<td>.67**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Max n-back level</td>
<td>.49*</td>
<td>.50*</td>
<td>-.12</td>
<td>-.27</td>
<td>.53*</td>
<td>.42</td>
<td>.10</td>
<td>.39</td>
<td>.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Average n-back level</td>
<td>.56*</td>
<td>.56*</td>
<td>-.29</td>
<td>-.26</td>
<td>.50*</td>
<td>.34</td>
<td>-.08</td>
<td>.35</td>
<td>.04</td>
<td>.95**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. FM score (control)</td>
<td>.17</td>
<td>-.53*</td>
<td>-.36</td>
<td>-.09</td>
<td>-.06</td>
<td>-.07</td>
<td>-.22</td>
<td>-.17</td>
<td>Na</td>
<td>Na</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. $P < .05* P < .001**$, two-tailed*
Coefficients for the positive relationships between \( n \)-back performance and SA were moderate, but not significant. Performance on the control feature match (FM) training was negatively associated with digit span score post-training, though not pre-training. These suggest links between performance on training and two of the transfer tasks, though not with Stroop latencies (RTs). Scores on TA did not correlate with digit span or Stroop performance, but there was a negative correlation between TA and IGT post-training which linked high anxiety with lower performance in IGT (H6). In further analyses, the total time spent in training over the 14 days did not correlate with performance on any transfer tasks, and, for those who completed valid training, there was a correlation between the maximum level of \( n \)-back achieved during training and cognitive stress post-training, \( r = -.512, P = .04 \) [\( n = 32 \)]. Cognitive stress is negatively scored, so higher stress was related to better training performance.

**Transfer tasks and trait anxiety.** In order to address the predictions that eWM \( n \)-back training would lead to gains in transfer tasks with a moderating effect of TA (H1-H4), firstly, hierarchical moderated regressions were performed on the gain scores (time 2 score - time 1 score) for the transfer tasks, but there was no interaction effect of training group and TA, and no effect of TA itself. Analysis of gain scores, however, can fail to take into account the starting differences in pre-test scores, where the availability of pre-test scores helps to reduce error variance, and they fail to make the most of pre-post randomised controlled test designs (Dimitrov & Rumrill, 2003). Therefore, a two-way RM MANCOVA was performed with time (pre- and post-training) and training group as IVs, trait anxiety as covariate, and DVs of digit span scores, Stroop incongruent trial latencies, and IGT scores. Box’s test, and Levene’s tests for the univariate effects, were all non-significant, meeting the test assumptions. Descriptive statistics for the tests are in Table 4.

There were no significant multivariate effects of time or training group, no covariate effect of trait anxiety, and no interactions that would show the differential effects of training (see Table 5 for the multivariate test results). Partialling out the effect of TA yielded no significance in the effect of training group. Examination of the univariate statistics revealed no significant effects of time on the dependent variables, though IGT scores between pre- and post-training showed a trend towards increasing \( F (1, 22) = 3.51, P = .07, \eta^2_p = .139 \). There were no univariate covariate effects of trait anxiety, and none of training group - though the effects of training group on Stroop overall was approaching significance (\( P = .07 \)), and is further investigated below. To test the assumption of homogeneity of regression slopes
required for ANCOVA models, the analysis was rerun with a custom model specifying the main effects and the interactions between training group and TA and the 3-way interaction. These were not significant, thereby meeting the assumption, and also justifying the use of a covariate model rather than moderated hierarchical regressions (Leppink, 2018).

Table 4

Descriptive statistics for control [n = 12] emotion working memory group [n = 14] on transfer tasks pre- and post-training

<table>
<thead>
<tr>
<th>Task</th>
<th>Training</th>
<th>Time 1 Pre-training</th>
<th>Time 2 Post-training</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M (SD)</td>
<td>95% CI</td>
</tr>
<tr>
<td>Digit span</td>
<td>Control</td>
<td>7 (1.41)</td>
<td>6.10, 7.9</td>
</tr>
<tr>
<td></td>
<td>eWM</td>
<td>7.07 (1.73)</td>
<td>6.07, 8.07</td>
</tr>
<tr>
<td>Stroop</td>
<td>Control</td>
<td>1106 (171)</td>
<td>997, 1214</td>
</tr>
<tr>
<td></td>
<td>eWM</td>
<td>1002 (164)</td>
<td>908, 1097</td>
</tr>
<tr>
<td>IGT</td>
<td>Control</td>
<td>2246 (898)</td>
<td>1675, 2816</td>
</tr>
<tr>
<td></td>
<td>eWM</td>
<td>2021 (972)</td>
<td>1460, 2583</td>
</tr>
</tbody>
</table>

*Note. Digit span = total scores; Stroop = averaged response latencies for incongruent trials in milliseconds; IGT = Iowa Gambling Task total score in dollars*

Table 5

Effects of the effects of time, training group and trait anxiety on transfers task scores

<table>
<thead>
<tr>
<th>Variables</th>
<th>Multivariate effects (all non-significant)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pillai’s Trace</td>
</tr>
<tr>
<td>Training group</td>
<td>.15</td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>.13</td>
</tr>
<tr>
<td>Time</td>
<td>.20</td>
</tr>
<tr>
<td>Time x Training group</td>
<td>.08</td>
</tr>
<tr>
<td>Time x TA</td>
<td>.13</td>
</tr>
<tr>
<td>Custom model interactions:</td>
<td></td>
</tr>
<tr>
<td>Training group x TA</td>
<td>.15</td>
</tr>
<tr>
<td>Time x Training group x TA</td>
<td>.05</td>
</tr>
</tbody>
</table>

The treatment of the between-subjects main effects in RM AN(C)OVA models is extremely conservative, as it includes pre-test scores not affected by the treatment effects.
(Dimitrov & Rumrill, 2003), and which is, in fact, identical to analysing gain scores. The analysis was therefore followed up with independent samples t tests with training group as IV and Stroop latencies (averaged incongruent trial RTs) pre- and post-training as DVs. There were no differences between the training groups on Stroop RTs at pre-training \( t(27) = .90, P = .38 \), as would be expected if groups were equal, but Stroop performance did differ significantly between training groups at post-training \( t(28) = 2.20, P = .04 \), with faster response times for the eWM group in comparison to the control group, \( M_{\text{experimental}} = 923, SD = 130, 95\% \text{ CI } [851, 995] \) and \( M_{\text{control}} = 1076, SD = 237, 95\% \text{ CI } [945, 1207] \). Therefore, we must reject H1, H3, and H4, and only H2 finds support. Stroop was the only area in which benefits of eWM training were seen, and TA did not show a moderating effect on any gains.

**Gender and IGT.** Independent samples t tests showed no significant differences in performance on the IGT at pre-training or post-training between men and women. However, a calculated gain score (post - pre scores) did differ between the sexes (see Table 6 for descriptive statistics), with the df and significance corrected for violation of the assumption of equality of error variances, \( t(31.93^{\text{corrected}}) = 2.44, P = .02 \). Women had higher gain scores than men, and so tended to improve more than men between pre- and post-training.

**Table 6**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Gender</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGT pre</td>
<td>Female</td>
<td>39</td>
<td>2312</td>
<td>951</td>
<td>2005, 2621</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>20</td>
<td>2282</td>
<td>916</td>
<td>1853, 2712</td>
</tr>
<tr>
<td>IGT post</td>
<td>Female</td>
<td>24</td>
<td>2468</td>
<td>1082</td>
<td>2012, 2712</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>11</td>
<td>1881</td>
<td>638</td>
<td>1453, 2311</td>
</tr>
<tr>
<td>IGT Gain</td>
<td>Female</td>
<td>24</td>
<td>335</td>
<td>919</td>
<td>52.62, 724</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>11</td>
<td>-270</td>
<td>495</td>
<td>-640, 100</td>
</tr>
</tbody>
</table>

**Stroop performance checks.** Response latencies, measured in milliseconds, were slightly faster at pre-training for congruent \( (M = 1107, SD = 338, 95\% \text{ CI } [1013, 1200]) \) than for incongruent trials \( (M = 1145, SD = 398, 95\% \text{ CI } [1145, 1035]) \) \( [n= 53] \) which should be expected if the test creates the desired affective interference, though the test figure was not significant in a paired samples t test, \( t(52) = 1.72, P = .091 \). At post-training the expected
differences between response latencies of congruent and incongruent trials was significant, respectively $M = 964$, $SD = 195$, 95% CI [895, 1033], and $M = 1014$, $SD = 309$, 95% CI [942, 1087], $t(32) = 3.77$, $P = .001$. Thus we can conclude that the incongruent trials provided affective interference and showed longer response latencies. Due to the presence of English as a second language participants in the sample (who may not have understood the meanings of all words), independent samples $t$ tests were performed on English first language vs other first language participants, and revealed no significant averaged Stroop RT differences either pre- ($P = .22$) or post-training ($P = .89$).

**Transfer tasks and COPSOQ items: cognitive stress and work demands.** Descriptive statistics and reliabilities for the COPSOQ items are presented in Table 7. Correlation analyses were performed among COPSOQ items and transfer task performance gains for the eWM training group, for which associations were predicted. High work demands did not correlate with gain scores in any of the transfer tasks, disconfirming H5. There were also no associations between cognitive stress (composite score) and gains in Stroop performance or with IGT (coefficients very close to zero), contrary to the hypothesis (H7).

<table>
<thead>
<tr>
<th>Items</th>
<th>$M$</th>
<th>$SD$</th>
<th>Cronbach’s $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive stress (composite) pre-training</td>
<td>12.45</td>
<td>3.84</td>
<td>.79</td>
</tr>
<tr>
<td>Cognitive stress (composite) post-training</td>
<td>14.72</td>
<td>3.14</td>
<td>.79</td>
</tr>
<tr>
<td>Work demands (composite) (14 items)</td>
<td>40.58</td>
<td>11.05</td>
<td>.91</td>
</tr>
<tr>
<td>Quantitative demands (3 items)</td>
<td>3.03</td>
<td>.95</td>
<td>.67</td>
</tr>
<tr>
<td>Cognitive demands (5 items)</td>
<td>2.73</td>
<td>.97</td>
<td>.83</td>
</tr>
<tr>
<td>Sensory demands (3 items)</td>
<td>2.29</td>
<td>1.06</td>
<td>.85</td>
</tr>
<tr>
<td>Demands for responsibility (3 items)</td>
<td>3.67</td>
<td>1.12</td>
<td>.68</td>
</tr>
</tbody>
</table>

Correlations on the whole sample showed IGT performance at post-training was significantly related to work demands, $r = .41$ $P = .02$. High values on the COPSOQ scales equate to low prevalence, so this suggests that those who performed better on the IGT at post-training had fewer work demands regardless of training group.
Anxiety and cognitive stress. The number of training sessions completed correlated negatively with trait anxiety, showing that anxious people tended to complete fewer sessions overall $r = -.32, P = .03, n = 49$. TA was correlated with reports of cognitive stress symptoms pre-training $r = -.53, P = .001$ and post-training $r = -.60, P = .001$, where the negative scoring of the cognitive stress shows that high anxiety was associated with increased reports of stress symptoms. This supports the prediction of a relationship between the two (H8). TA was also significantly associated with quantitative work demands (i.e. work load) $r = -.41, P = .001$, though not with other work demand scores. The subscale of quantitative demands (negative scale) also correlated with SA at pre-training $r = -.31, P = .02$, and cognitive stress at pre-training $r = .34, P = .01$, showing high workload and time demands linked to high anxiety and stress, though the composite work demands score did not show such relationships.

Table 8

Descriptive statistics for training groups on state anxiety and cognitive stress.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Training</th>
<th>Time 1 Pre-training</th>
<th>Time 2 Post-training</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M (SD)</td>
<td>95% CI</td>
</tr>
<tr>
<td>SA</td>
<td>Control</td>
<td>33.13 (10.11)</td>
<td>27.53, 38.73</td>
</tr>
<tr>
<td></td>
<td>eWM</td>
<td>33.25 (9.80)</td>
<td>28.03, 38.47</td>
</tr>
</tbody>
</table>

Note. High cognitive stress scores equate to low reports (negative scale). There were no significant univariate effects of any kind on SA scores.

A two-factor mixed MANOVA with time (pre vs. post) and training group as IVs and SA and cognitive stress scores as DVs, revealed a significant multivariate main effect of time Pillai’s Trace = .99, $F (2, 28) = 6.47, P = .005, \eta^2_p = .316$, but no main effect of training group, Pillai’s Trace = .004, $F (2, 28) = .06, P = .94$, and no interaction effect, Pillai’s Trace = .002, $F (2, 28) = .03, P = .97$. Univariate statistics showed a significant main effect of time on cognitive stress, $F (1, 29) = 7.59, P = .01, \eta^2_p = 207$, with stress scores decreasing over time overall (scale is negatively scored), but no interaction that would imply differential
improvement for training groups, and no main effect of training. The means are in the opposite direction on SA scores, as they increase over time, rather than decrease, showing the opposite pattern to stress scores (which are negatively scored) (see Table 8). The lack of interaction effects does not support the prediction that n-back training would decrease anxiety and stress scores in comparison to the control training group (H9). Note that a significant Box’s test of equality of covariance matrices ($P = .01$) calls the reliability of the multivariate tests into question.

**Work-related anxiety.** A coding variable was calculated based upon whether participants rated their current anxiety level as moderately to highly related to their work (3 and above on the 5-point scale). And this was combined with median split scores on SA to create an IV of individuals who were highly anxious about their work, to analyse the effects on training performance, and on the pre- and post-training training scores. The average n-back performance of participants who were anxious about their work at pre-training was poorer than that of those scoring low in work-related anxiety $t (16) = 2.15, P = .047$, in the high anxiety group $M = 3.30, SD = 1.44, 95\% CI [2.27, 4.33] n = 10$, and in the low anxiety group $M = 4.70, SD = 1.26, 95\% CI [3.64, 5.76] n = 8$. So those who were currently highly anxious about work obtained a lower level of n-back on average over the 14 days. They also did fewer training sessions in the subsequent training $t (27) = 2.14, P = .04$, in the high anxious group $M = 9.19, SD = 5.65, 95\% CI [6.18, 9.19] n = 13$, and in the low anxious group $M = 13.15, SD = 3.93, 95\% CI [10.78, 15.53] n = 16$. Work-related anxiety scores at post-training were linked to higher subjective reports of symptoms of cognitive stress over the past 4 weeks at post-training, $t (17) = 3.33, P = .004$, for the highly anxious group $M = 12.30, SD = 2.71, 95\% CI [10.36, 14.24] n = 10$, and for the low work anxious group $M = 16.33, SD = 2.55, 95\% CI [14.37, 18.29] n = 9$, where low scores = more cognitive symptoms. Change scores (time 2 - time 1) in work related state anxiety did not differ significantly between control [$M = -.07, SD = 1.28$] and experimental groups [$M = -.38, SD = 1.31$] (where the scale is 1-5), and so did not decrease as a result of training $t (29) = .66, P = .51$.

**Students and professionals.** Full time students ($n = 23$) and those in full time employment ($n = 20$) were compared on transfer task performance, anxiety scores, and on subjective ratings of work demands and cognitive stress symptoms. Those in full time studies and full time employment did not differ in performance on transfer tasks, nor in their training performance. Trait anxiety scores did not differ between the two groups ($P = .33$), showing comparable levels of dispositional anxiety. Students and professionals did differ on work demands (composite score) $t (41) = 2.70, P = .01$, with students less likely to report demands...
from their studies \([M = 40.39, SD = 7.6, 95\% \text{ CI } [37.1, 40.68]\) than professionals did from their work \(M = 34.6, SD = 6.25, 95\% \text{ CI } [31.67, 37.52]\). In terms of the subscale scores on demands, students reported cognitive demands less often \((P = .01)\), and fewer demands for responsibility \((P = .01)\). Reports of symptoms of cognitive stress did not differ between students and professionals, either at pre- \((P = .53)\) or post-training \((P = .25)\). Students were more likely \((t (41) = 3.02, P = .01)\) to attribute their current level of anxiety at pre-training to work/studies, \(M = 3.61, SD = 1.34, 95\% \text{ CI } [3.03, 4.19]\), than those in full time employment \(M = 2.40, SD = 1.27, 95\% \text{ CI } [1.80, 3]\). There were no other significant differences in demands and stress scores.

### Table 9

**Post-training evaluation questions (scale 1-7, Not at all—Very much so) for each training group: descriptive statistics**

<table>
<thead>
<tr>
<th>Item</th>
<th>Group</th>
<th>(M)</th>
<th>(SD)</th>
<th>(n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I feel the brain training improved my mental sharpness</td>
<td>FM Control</td>
<td>3.41</td>
<td>1.87</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>eWM</td>
<td>3.43</td>
<td>1.79</td>
<td>14</td>
</tr>
<tr>
<td>I feel the brain training improved my memory</td>
<td>FM Control</td>
<td>3.24</td>
<td>1.92</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>eWM</td>
<td>3.29</td>
<td>1.73</td>
<td>14</td>
</tr>
<tr>
<td>I feel the brain training improved my concentration</td>
<td>FM Control</td>
<td>3.88</td>
<td>2.09</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>eWM</td>
<td>3.43</td>
<td>1.99</td>
<td>14</td>
</tr>
<tr>
<td>I feel the brain training improved my ability to control my emotions</td>
<td>FM Control</td>
<td>2.53</td>
<td>1.59</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>eWM</td>
<td>2.43</td>
<td>1.60</td>
<td>14</td>
</tr>
<tr>
<td>I feel the brain training helped me cope with work or study stress</td>
<td>FM Control</td>
<td>2.94</td>
<td>1.68</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>eWM</td>
<td>2.57</td>
<td>1.74</td>
<td>14</td>
</tr>
</tbody>
</table>

**Subjective evaluation of the effects of training.** A one-factor MANOVA with training group (eWM vs control) as IV and scores on the brain training evaluation items as DVs showed no multivariate \(F (5, 25) = .68, P = .65\), and no univariate effects. Neither group differed in their subjective evaluation of the benefits of the training (see Table 9 for descriptive statistics).

**Participant insight into treatment groups.** Pearson’s chi-squared test was used to analyse if there was an association between the participants’ guess about which training group they were in and their actual training group allocation (Table 10). The was no
significant association between the two, $\chi^2(2) = .76, P = .68$, suggesting that overall participants did not correctly guess which training group they had been in.

Table 10

*Chi-squared results showing the association between training group allocation and participant’s guess*

<table>
<thead>
<tr>
<th>Group guess</th>
<th>Exp. group</th>
<th>Control group</th>
<th>Not sure</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>Observed</td>
<td>8</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Expected</td>
<td>9.0</td>
<td>5.8</td>
<td>3.2</td>
</tr>
<tr>
<td>Experimental</td>
<td>Observed</td>
<td>9</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Expected</td>
<td>8.0</td>
<td>5.2</td>
<td>2.8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>Observed</td>
<td>17</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Expected</td>
<td>17.0</td>
<td>11.0</td>
<td>6.0</td>
</tr>
</tbody>
</table>

**Discussion**

The first aim of the current study was to replicate findings showing benefits of emotional working memory training in transfer (i.e. untrained) cognitive performance tasks, that were purposely selected to provide further support for theoretical models of the cognitive and neural substrates of decision-making under stress and anxiety. The second aim was to link executive performance to work demands, stress and anxiety. The goal of the research was to find support for eWM training as a potential cost-effective intervention to reduce work stress and anxiety, with concomitant benefits for performance and well-being in organizations. The first aim was only partially achieved, and neither digit span as a near transfer task of working memory capacity, nor the far transfer task of decision-making under uncertainty, the Iowa Gambling Task, improved after eWM training relative to an active control group. The medium transfer task, emotional Stroop, a gold-standard measure of executive control (Schweizer et al., 2011), did show greater gains after training compared to the control treatment group. The anticipated moderating interactions of trait anxiety on gains that should be expected from the literature were also not observed. The second study aim received similarly weak support. The eWM training did not reduce state anxiety or cognitive stress symptoms relative to the control training. Relationships were found between stress, and between work demands and performance on the eWM training, and this supports the idea that stress has demonstrable effects on cognitive functioning. With these aims unachieved, the
goal of the research is currently obscured, though it may be too early yet to abandon it. Some interesting patterns that confirm previous work were uncovered, and results are discussed in detail below.

Digit span did not improve between pre- and post-training for either group, therefore rejecting H1. There was very high stability in digit span scores overall though, and it may be that forward digit span was too limited a measure to detect training gains in working memory capacity. However, higher performance on the \(n\)-back training was related to higher scores on digit span (both pre- and post-training), suggesting a close link between the two, which would be predicted for the near transfer task. The fact that span was correlated with performance on \(n\)-back pre-training, in addition to post-training, does suggest the causal link, if any, is in the direction of working memory capacity leading to better performance on \(n\)-back. The same relationships between \(n\)-back performance and IGT were also observed. This is consistent with the proposal that the tasks rely on shared cognitive processes, specifically those underlying executive function, which is supported by the ample literature on these tasks, and also on the training literature that show transfer gains. A simpler explanation might be that participants who tried hard on the training also tried harder and did better on the transfer tasks. Although this latter explanation does not seem consistent with the finding that those who performed less well in the feature match control group were more likely to show higher gains in digit span. Clearly effort in that task was not related to performance in the transfer tasks in this case. This curious finding in itself is tricky to explain, unless this was down to boredom with the FM training and therefore more effort was put into the post-training tests that posed more of a challenge.

If the aim is to improve performance and well-being under work stress, then the intervention must show clear gains on a task of emotion regulation. The Stroop results were therefore crucial to interpretation of the findings and their implications for organizations. Overall, Stroop latencies did appear to be longer for the incongruent trials, suggesting that the emotional variant of the Stroop test devised by Schweizer and colleagues was successful in inducing interference between the valence of the word presented (and which must be responded to) and the valence of the facial expression on which the word is superimposed and which must be ignored. Because the sample contained some participants for whom English has not a first language, this was also analysed for potential effects, but was unrelated to Stroop performance. Though gains were not seen overall in the multivariate analysis, performance on the incongruent trials did change after emotional dual \(n\)-back training relative to the active “placebo” control group, supporting H2. This is consistent with Schweizer et
al.’s (2011) findings, explained by improvements in the processing of interference of emotion-laden stimuli, and with other research showing medium to far transfer gains from n-back training (Au et al., 2015), though it should also be considered that it could represent regression to the mean (though there is no suggestion of this in the test results of other transfer tasks, which were not significantly different). This finding is important in suggesting that the theoretical foundation of any gains in eWM training lies in improved executive function, as Stroop is a ‘gold-standard’ measure of this, and it is further implied, from the affective stimuli used in the training and emotional variant of the Stroop task itself, that this is linked to better emotion regulation (Schweizer et al, 2011). However, when this finding is considered in tandem with the lack of a moderating or covariate effect of trait anxiety in the present study, and rejection of H4, it cannot be said to support the involvement of better emotion regulation definitively. The construct of TA is well-established, and high TA individuals have faulty emotion regulation as evidenced in behavioural and brain imaging measures. The STAI measure used in the present study has high validity and reliability. Therefore, from the present findings is cannot be concluded that the affective variant of dual n-back improves emotion regulation, though it may well act to improve executive processing generally. This could instead show gains though enhancing general resource efficiency, and the management of cognitive load. The inclusion of a non-affective version of Stroop alongside the affective one in a future study would permit examination of this, as would fMRI measures as in Schweizer et al. (2013).

One alternative explanation for the Stroop gains might lie in the nature of the control group used. As several authors have noted (Melby-Lervåg et al., 2016; Redick & Lindsey, 2013; Redick, 2015), detecting transfer gains after n-back training seems sensitive to the type of comparison/control group, with n-back gains less likely from comparisons with active control groups than passive. The present study used an active group, and, further, corrected for a potential confound in Schweizer et al.’s published studies which did not use face stimuli in the control training, where effects could be down to simple exposure to faces in the n-back group. The control task used here was that devised by Schweizer in her latest, unpublished work, and used a face match task to address the potential confound. However, it remains that there is still a potential confound with the control training. The feature match training requires participants to scan two arrays of faces with different individuals and facial expressions, and to respond if the panel is a match or mismatch, in a visual scanning task. Panels can be a mismatch not only if a different individual face is presented in some position on the second panel, but also if the same individual face appears but with a different facial
expression. It therefore requires close attention to the face and the facial expression. The affective $n$-back training, on the contrary, requires participants to ignore both the faces and facial expressions. As the Stroop task also requires one to ignore the interfering faces, the question is: does the affective, facial stimuli version of $n$-back produce gains in emotional Stroop performance merely through training participants to ignore faces? The present findings could be explained by this, and not necessarily through enhancing executive functioning per se. Stroop cannot be truly said to be a transfer task if the demands match too closely the training. The control training therefore requires further modification in future studies, to present a task that uses the same emotional faces but that does not require attention to be paid to those faces, while still making minimal demands on memory and executive processes. Only then can it be concluded definitively, from behavioural studies like the current one at least, that emotional $n$-back improves executive function. Further evidence using more elaborate behavioural measures of emotion regulation, and preferably brain imaging, are required to make conclusions about the affective version of the $n$-back task on emotion regulation.

The were no gains in IGT performance for eWM training in comparison to control training, rejecting H3. Performance on $n$-back itself was related to better IGT performance – pre- and post-training, consistent with some shared cognitive processes. Clearly there was a great deal of variability in changes in IGT performance, and large standard deviations. It may therefore not be an adequate test of training gains. IGT could well be sensitive to repeated testing (though there is scant evidence of substantial practice effects in the data set), or that it is affected by factors like mood, time of day, environmental distractions, and which is indeed a problem in interpreting all the pre- and post-training data from the study. Indeed, the gender difference found in previous research (den Bos, Homberg, & de Visser, 2013), where women are less likely to choose the advantageous decks, was not supported here. The current sample was predominantly female, though less so in the pre-training sample, and only 11 males completed post-training in total and 8 valid completers. IGT gains were actually lower in males than in females, in the valid training sample with one female outlier omitted, the IGT the difference was $585 in favour of women. It may be that women showed more conscientiousness in completing the training and engaging with the post-training IGT task. Low motivation in the task may have been a problem in general, where the ‘monetary gains’ were hypothetical. Indeed, research has shown IGT scores are more variable using facsimile money rather than real money (Bowman & Turnbull, 2003). If there is no arousal associated
with the bad decks, then participants will not be guided by the somatic feedback to avoid those decks.

Work demands did not correlate with performance in the transfer tasks in the eWM group, showing no support for H5. There was a trend for reports of low work demands to correlate with better n-back performance, which suggests that those who perhaps have a lower cognitive burden in their everyday work or studies were able to engage more in the training task. Overall, IGT performance at post-training was linked to low work demands. It may be that cognitive exhaustion in the high job demands participants prevented them from engaging fully with the training and transfer tasks. The composite score of work demands did not show any relationships to stress or anxiety, however, the subscale of quantitative demands, i.e. workload and time demands, did show expected associations with state and trait anxiety and reports of cognitive stress, supporting the idea of a link between work demands and stress and anxiety.

Performance on the IGT, measured in gain scores between pre- and post-training performance, was negatively associated with trait anxiety, suggesting that the trait anxious participants in the study showed less gain in performance after training overall, and IGT scores at post training were also negatively correlated with TA, partially supporting H6 (though no other relationships to transfer task performance were found with TA). This confirms previous work (Miu et. Al., 2008) on IGT performance and TA. The results are consistent with the finding that dispositionally anxious individuals show impaired processing of risk in this task, potentially because of faulty modulation of feedback from somatic markers. High TA scores were also significantly associated with subjective reports of cognitive stress in the last month, including symptoms such as problems with memory, decision-making, thinking clearly, and concentration, supportive of the idea that anxiety interferences with executive function, and supporting H8.

Neither cognitive stress, nor state anxiety showed any clear associations with transfer task performance, though SA at pre-training was significantly associated with longer response latencies to Stroop incongruent trials, and which does suggest anxious people had more difficulty avoiding affective interference in partial support of H7. As mentioned, trait anxiety did show an association with reports of cognitive stress symptoms, both pre- and post-training, and similarly, reports of high cognitive stress were significantly correlated with state anxiety scores. These results together suggest that anxious individuals were more likely to perceive impairments in their own cognitive performance over the previous month. Though measured through subjective means, this is an indication that work stress and anxiety
is linked to cognitive failures in attention, memory and thinking. It is therefore even more disappointing that the training had no effects on stress, at least as measured in the study, but it does highlight the importance of this line of intervention.

State anxiety scores did not decrease for the eWM group in comparison to the control group between pre- and post-training and neither did cognitive stress scores, and therefore eWM did not have demonstrable effects on stress and anxiety and H9 must be rejected. Cognitive stress decreased over time across the whole sample, which may well be down to a placebo effect, as cognitive stress is measured as subjective reports of problems with memory, concentration, decision-making, and thinking, and which participants’ insight should perhaps obviously connect to training. As these scores are based on subjective judgements of cognitive symptoms in the past 4 weeks, they may not be an accurate or reliable reflection of actual cognitive state. Future studies might examine this with supplementary objective measures. Life events may intercede to affect scores on pre- and post-training tests, especially influencing mood and state anxiety, through cognitive stress symptoms perhaps less so. These problems are compounded in a small sample, where such error variance will have a greater influence on the ability to find effects.

There was a non-significant trend for non-completions to have higher cognitive stress symptoms pre-training (self-reported problems with memory, concentration, taking decisions and thinking clearly). Those for whom training might have the most benefits were unfortunately slightly less likely to complete. This result was mirrored by that on TA, and those who failed to complete a valid training programme had higher TA and SA pre-training. This might have affected statistical power of tests to detect the predicted interaction that expected this group to benefit the most in gains from n-back. For those who did complete the study, those achieving a higher level of n-back during the training reported higher cognitive stress post-training. This is a curious finding that seems somewhat contradictory, given their performance on n-back suggests good working memory and attention rather than the opposite. Perhaps those who performed well on the task were more disappointed with the demonstrable outcomes of the training, had higher expectations, or perhaps subjective ratings of cognitive stress symptoms are only weakly related to objective condition.

People who were coded as currently highly anxious about work, obtained a lower level of n-back performance on average over the 14 days’ training. This suggests that work-related anxiety specifically may impair working memory performance, though it could also be explained by these participants being less engaged in the training task than low work anxious individuals. Work-related anxiety scores at post-training were linked to higher
subjective reports of cognitive stress symptoms over the past 4 weeks at post-training, i.e. problems with memory, concentration, thinking and decision-making. These findings together support a body of work that shows work-related anxiety has deleterious effects on general aspects of everyday life outside of the work sphere, and with effects extending to cognitive functioning (Ganster & Rosen, 2013), though it must be noted that the sample size was very small in these comparisons, making generalisations difficult. Higher reports of symptoms associated with poor cognitive functioning perhaps provide further impetus to targeted interventions that focus on these aspects of work-related stress, and not simply emotional coping or support mechanisms, though emotion and cognition is inextricably linked.

Students and those in full-time employment did not differ in their performance on either training or transfer tasks. The two groups did not differ in terms of reports of cognitive stress, and showed similar levels of anxiety, though work demands, did vary between the two groups, perhaps unsurprisingly. The latter does not support other research suggesting that students were more prone to anxiety than the general population (Wöfel et al., 2016), though it may be that the sample of students in the study were, through their observed low demand scores, under less stress than the students in the Wöfel et al study. Many of the students in the current study were international students, and thus the sample could provide a different demographic. It does suggest that students and those in paid work, are, on the whole, quite similar in both performance and in cognitive stress.

In sum, there is support in the current data set for the idea that anxiety in general, work-related anxiety, and stress symptoms all link to poor cognitive performance, and also to performance in the eWM training itself, though there is only partial support for a link to performance in the transfer tasks. Participants subjective evaluations of the benefits of training also did not differ between the eWM and placebo groups. Work demands did not show relationships to training performance or training gains, and so eWM did not work better with participants who have high work demands, and did not appear to augment the work demands faced in their daily work or studies. The intervention did not lead to improvements in cognitive stress scores, or have more benefits for participants who experience high work demands. Anxiety scores were significantly related to work demands and reports of cognitive stress, which corroborates theoretical models of workplace stress, such as Karasek’s demand-control-support model, and the many studies showing links between demand factors like workload and cognitive strain (Hambrick et al, 2005). Some work demands might well be beneficial to cognition (Then et al., 2014), and extend cognitive capacity, though others
might work detrimentally. Indeed, analysis of the subscale scores on work demands showed that quantitative work demands, work load and time demands, were linked to stress and anxiety. There will also be individual differences, and demands may provide challenge for some individuals and threat for others. Further work would need to explore individual appraisal processes (Lazarus & Folkman, 1984), coping styles and self-efficacy to shed further light on these links.

There are a number of limitations in methodology and sampling that could also explain the null findings generally. The small sample size of the present study and lack of power to detect differences make conclusions uncertain and effects hard to generalise. The analysis of the multiple factors measured in the current study would perhaps benefit from structural equation modelling (SEM) to simplify the complex structure of relationships observed. A key factor in explaining the results is perhaps the lack of control in the pre- and post-training testing situation, as this was done online. This alone may account for the inability to find the predicted effects in pre- and post-training tests. Schweizer’s studies conducted pre- and post-training in the lab, keeping time of day constant, and permitting testing under controlled conditions. Resource restrictions such as lack of time and funds in the current study prevented this. The practicalities of obtaining participants in such a short time, with no payment system, necessitated recruitment online, and the sample was distributed over a wide geographical area and across national boundaries.

Indeed, if any one lesson can be learned from the present research it is perhaps on the nature of an online study using a pre- and post-test design. A better resourced study, with payment system in place for participation, which is common in the many training studies, could overcome this limitation. It is clear from personal communications with some participants, who emailed feedback at the end of the study, that conditions of post-training testing were not optimal for them, and this will have affected the results, especially with such a small sample. IGT specifically has been found to be affected by negative mood (Suhr & Tsanadis, 2007). It was hoped that by measuring state anxiety, the influence of current state could be accounted for, however, it may be that wider measurers of mood and state were required beyond SA to control for the potential influence of extraneous variables on pre- and post-training scores. There are certainly other dispositional factors that have been found to influence performance on the tasks that were not measured, including intellectual ability and fluid intelligence (Suhr & Hammers, 2010), addictions (Bechara & Damasio, 2002), and impulsivity and sensation seeking (Buelow & Suhr, 2009) and which could in particular have impacted upon tasks of executive function, and IGT performance specifically.
Conclusions. To my knowledge, no published replications of Schweizer et al.’s eWM variant of \( n \)-back exist outside of that lab group. This study is therefore important in attempting to replicate the findings with a more diverse sample. The observed links between work demands and cognitive stress, and work-related anxiety and the \( n \)-back training, provide impetus to future research, and point the way for improvements such as in removing potential confounds of the control training group. It also represents an initial foray into the influence of WM training on decision-making, and its relationship to individual differences, and which in future research can, in turn, be linked more directly to the workplace with samples drawn from highly demanding professions such as the police, air traffic controllers, and emergency room professionals. Further studies applying WM training to organizations and stress would also benefit from more detailed measures of workplace stress and types of stressor, and through measuring aspects of job control and support, tapping into individual appraisal processes, self-efficacy and coping styles.

It is, then, premature to conclude that training using emotional \( n \)-back, or using working memory training generally, is of no importance to organizational researchers. The current study provides a lesson in some of the pitfalls of conducting cognitive training interventions and points the way to improvements in the design and implementation of future ones. The relationships that were found in the data, between work demands, cognitive stress, state and trait anxiety, and \( n \)-back training performance itself, suggest the theoretical underpinnings of the current study are sound. Future studies that overcome resource and methodological limitations are required to achieve the laudable goal of improving organizational performance and well-being through an implicit cognitive intervention to target executive functioning.
References


Johnson, D. R., & Gronlund, S. D. (2009). Individuals lower in working memory capacity are particularly vulnerable to anxiety’s disruptive effect on performance. *Anxiety, Stress, & Coping, 22*, 201–213.


Appendices

A. COPSOQ items
Five-point scales: 1 Always, 5 Never/Hardly ever

Work demands:

Quantitative demands
Do you have to work very fast?
Is your workload unevenly distributed so it piles up?
How often do you not have time to complete all your work tasks?

Cognitive demands
Do you have to keep your eyes on lots of things while you work?
Does your work require that you remember a lot of things?
Does your work demand that you are good at coming up with new ideas?
Does your work require you to make quick decisions?
Does your work require you to make difficult decisions?

Sensory demands
Does your work demand a great deal of concentration?
Does your work demand your constant attention?
Does your work require a high level of precision?

Demands for responsibility
Could it injure other people if you make mistakes in your work?
Could it cause financial losses if you make mistakes in your work?
Does your work affect the well-being of others?

Cognitive stress:

How much of the time during the past 4 weeks have you...
...had problems concentrating?
...had difficulty in taking decisions?
...had difficulty with remembering?
...found it difficult to think clearly?

B. Brain training evaluation questions (post training)

Indicate the extent to which you agree with the following statements, scale 1-7, not at all – very much so

I feel the brain training improved my mental sharpness
I feel the brain training improved my memory
I feel the brain training improved my concentration
I feel the brain training improved my ability to control my emotions
I feel the brain training helped me cope with work or study stress