2012

Improving memory using N-back training

Paul Beavon
Edith Cowan University

2012

Improving memory using N-back training

Paul Beavon
Edith Cowan University

Use of Thesis

Signed [redacted]

Dated 28 October, 2012______________________________

Improving Memory Using *N*-Back Training

Paul Beavon

A report submitted in Partial Fulfilment of the Requirements for the Award of Bachelor of Science (Psychology) Honours, Faculty of Computing, Health and Science,

Edith Cowan University.

Submitted (October 2012)

I declare that this written assignment is my own work and does not include:

(i) material from published sources

used without proper acknowledgement; or

(ii) material copied from the work of other students.

Signature: [redacted]

Date: 28 October, 2012______________________________

Improving Memory Using *N*-Back Training

Abstract

Investigations into *n*-back training and near transfer to short-term memory (STM) and working memory (WM) have realised inconsistent results. A significant transfer to STM was reported using dual *n*-back training (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008). However, the majority of studies have found no significant transfer to WM as operationalised by complex span tasks using either single or dual *n*-back training. The current study examined the single *n*-back task and near transfer to STM and WM as operationalised by the Woodcock-Johnson III Tests of Cognitive Abilities (Mather & Woodcock-Johnson, 2001). Forty-seven participants were divided into experimental treatment (n = 26) and active control (n = 21) groups; and engaged in 20 daily, 20-minute training sessions over a 30-day period using either a single *n*-back task, or a combination of two general knowledge tasks respectively. STM and WM psychometric tests were administered before and after the 30-day training process. No significant difference was found between pre- and post-training STM or WM scores, indicating both constructs were unlikely near transfer mechanisms for single *n*-back training. There was concern that the non-significant WM finding may have been confounded as there is evidence to suggest that the single *n*-back task and one of the active control group tasks both relied on recognition for resolution. The small effect size associated with single *n*-back transfer to STM implied that this outcome was independent of the active control group. Furthermore, the non-significant result for STM suggests that single and dual *n*-back tasks differ in their transfer properties.

Paul Beavon

Dr Ken Robinson

Dr Ricks Allan

COPYRIGHT AND ACCESS DECLARATION

I certify that this thesis does not, to the best of my knowledge and belief:

(i) incorporate without acknowledgement any material previously submitted for a degree or diploma in any institution of higher education;

(ii) contain any material previously published or written by another person except where due reference is made in the text; or

(iii) contain any defamatory material.

Signed [redacted]

Dated 28 October, 2012

Acknowledgments

Firstly, I would like to thank Dr Ken Robinson and Dr Ricks Allan for their encouragement, interest, insight and support throughout the research process. I would also like to thank Walter Howe, the Australian distributor of the Woodcock Johnson III Tests of Cognitive Abilities for his technical support and for providing softcopy templates of the required test forms free of charge. Finally, I would like to thank Allison Prandl who conducted a parallel research project in which the data collection process was shared. Her persistence, drive and commitment ensured we had sufficient participants to make both our projects viable.

Table of Contents

List of Tables

List of Figures

Improving Memory Using *N*-Back Training

Does *n*-back training mediate working memory (WM) and short-term memory (STM)? Jaeggi, Buschkuehl, Jonides, and Perrig (2008) maintained that their reported improvement in fluid intelligence (Gf) was realised through an improvement in WM, despite finding no improvement in WM as measured by reading span (Daneman & Carpenter, 1980). This was theoretically evidenced through the existence of a similar hypothesised capacity constraint identified in both WM and reasoning (Halford, Cowan, & Andrews, 2007); the concordance of central nervous system usage between WM and Gf tasks (Kane & Engle, 2002); and the substantial variance shared between working memory capacity (WMC) and Gf (Ackerman, Beier, & Boyle, 2005).

However, Jaeggi, Studer-Luethi, et al. (2010) modified their stance after a failure to find a significant improvement in operations span (Turner & Engle, 1989) with either single or dual *n*-back training. They advanced the hypothesis that the *n*-back task relied on recognition, rather than active recall, a process necessary for operations span. Contrary to this supposition, Jaeggi et al. (2008) found a significant improvement in STM using digit span, a measure that requires active recall. However, this has not been further investigated.

The near transfer mechanisms of the *n*-back task are little understood (Shipstead, Redick, & Engle, 2012). Ongoing investigations of *n*-back based cognitive training regimes have produced inconsistent findings concerning the measures of WM. As previously stated, Jaeggi et al. (2008) found no improvement in reading span using dual *n*-back training; and Jaeggi, Studer-Luethi, et al. (2010) found no improvement in operations span with either single or dual *n*-back training. Similarly, after an extended training duration on two different spatial *n*-back tasks Li et al. (2008) found no improvement in operations or rotational span. However, in an ongoing study, Seidler et al. (2010) found a significant improvement in operations span, using dual *n*-back training. Finally, in a broad cognitive training study that included a spatial *n*-back training task, Schmiedek, Lövdén, and Lindenberger (2010) found no improvement in reading and counting span for younger and older adults, however a significant improvement in rotational span was realised for older adults.

The possible explanations for these inconsistencies are varied. Firstly, the *n*-back application is not a standardised instrument, and was likely to differ in functionality between studies. Secondly, the *n*-back modality was not consistent, with some studies having used single *n*-back which relies on visual stimuli (Jaeggi, Studer-Luethi, et al., 2010; Li et al., 2008; Schmiedek et al., 2010), whilst others have used dual *n*-back which combines both visual and auditory stimuli (Jaeggi et al., 2008; Jaeggi, Studer-Luethi, et al., 2010; Seidler et al., 2010). Further, single *n*-back stimuli included both numbers (Li et al., 2008) and spatial location (Jaeggi, Studer-Luethi, et al., 2010; Li et al., 2008; Schmiedek et al., 2010). Thirdly, the memory loading or *n* was either manually set between two and four (Li et al., 2008; Schmiedek et al., 2010), or allowed to vary with participant ability (Jaeggi et al., 2008; Jaeggi, Studer-Luethi, et al., 2010; Seidler et al., 2010). Fourthly, training duration has ranged between 17 (Seidler et al., 2010) and 197 days (Schmiedek et al., 2010), with the most studies averaging around 20 days (Jaeggi et al., 2008; Jaeggi, Studer-Luethi, et al., 2010; Seidler et al., 2010). Finally, experimental active control groups were not always used, and therefore some studies were exposed to potential internal validity problems (Jaeggi et al., 2008; Jaeggi, Studer-Luethi, et al., 2010).

Much of the previous *n*-back near transfer research has focused on mechanisms that would facilitate far transfer to Gf through known pathways identified using latent variable analysis (Conway, Kane, & Engle, 2003) and theoretical suppositions. This has resulted in a less than systematic approach that has been reliant on manifest variable relationships. A feature of the current study is that it was theoretically based on the Cattell-Horn-Carroll

(CHC) theory factorial model of intelligence that is operationalised by the Woodcock-Johnson III Tests of Cognitive Abilities (WJ-III; Mather & Woodcock-Johnson, 2001).

The aim of the current study was to investigate the near transfer mechanisms of single *n*-back training. The constructs of interest were WM and STM. To date, the majority of evidence implies WM as operationalised by complex span tasks is unaffected by single and dual *n*-back training, however broader measures of WM are required before eliminating this construct as a near transfer mechanism (Shipstead et al., 2012; Sternberg, 2008). Also, the significant increase in digit span through a dual *n*-back training regime (Jaeggi et al., 2008) provides support for the near transfer to STM using single *n*-back training.

N-Back Tasks, Training and Transfer

The *n*-back task. The *n*-back task is a computer based cognitive activity (Gray, Chabris, & Braver, 2003; Jaeggi et al., 2008; Jaeggi, Buschkuehl, Jonides, & Shah, 2011; Jaeggi, Studer-Luethi, et al., 2010) that presents a finite set of stimuli in a continuous stream, where the participant is required to respond to those stimulus that match stimuli delivered *n* positions previously. Stimuli are usually visual or auditory. Visual stimuli are a characteristic of the single *n*-back tasks and include shapes, images, letters, words, and numbers either displayed individually or located in spatial arrays. The single *n*-back task can be modified to dual *n*-back with the addition of auditory stimuli such as letters or words.

N-back tasks have acquired their own nomenclature. Each trial is called a *block* and consists of a set number of stimuli. The time interval between the presentations of consecutive stimuli within a block is the *inter-stimulus interval*. A stimulus that meets the *n*-back criteria for a match is called a *target* or *control target*. Stimuli that are targets but presented at position *n*+1 or *n*-1 are called *lures* or *lure foils*. All other stimuli presented within a block are described as *foils* or *control foils*. *Memory load* or *load* directly reflects the value of *n*.

***N*-back training and transfer**. Using a dual *n*-back task Jaeggi et al. (2008) established that significant near transfer to STM had occurred after 19 consecutive days of training. However, there was no significant change to reading span within this timeframe. Each training day comprised 20 blocks that were required to be completed in a single contiguous 25-minute session. In a more recent paper Jaeggi, Studer-Luethi, et al. (2010) compared the transfer efficacy between dual and single *n*-back training tasks, finding that neither task was effective in improving operations span after 20 consecutive training days, completing 15 blocks each daily session. However, Seidler et al. (2010), reported transfer to operations span using dual *n*-back training over a 17 to 25 day duration, with participants completing 20 blocks per session.

Definition and Operationalisation of STM and WM Constructs

STM. STM refers to both a theoretical storage system, the precursor to WM within the modal model (Atkinson & Shiffrin, 1968); and a simple finite capacity temporary storage facility, where information is consciously held for up to 30 seconds (Engle, Tuholski, Laughlin, & Conway, 1999). The capacity of STM is typically measured using simple or STM span tasks such as digit span (Conway et al., 2005), which requires the recall of a string of single digit numbers in correct serial order that was previously presented. Both forward and backward span are considered STM measures for adults (Conway et al., 2005; St Clair-Thompson, 2009); however they are also considered measures of WM (Mather & Woodcock-Johnson, 2001; Unsworth & Engle, 2007b; Wechsler, Coalson, & Raiford, 2008).

Jaeggi et al. (2008) measured STM using the digit span subtest of the Hamburg-Wechsler Adult Intelligence Scale-Revised, which is considered a measure of short term verbal memory (Tewes, 1991). The digit span subtest requires the subject to repeat in serial (forward digit span) and in reverse serial order (backward digit span), up to nine digits read by the administrator. The sum of the forward and backward components provides a measure of performance (Spreen & Strauss, 1998).

WM. Perhaps the most recognised WM framework and the successor to the modal model, is the multi-component model of WM developed by Baddeley and Hitch (1974). The authors described WM as a "limited capacity cognitive system allowing for the temporary storage and manipulation of information for complex tasks such as comprehension, learning and reasoning" (Baddeley, 2000, p. 418). The three-component model comprises a domain general attentional controller, the *central executive*; which is assisted by two STM domain specific slave systems, the *phonological loop* and *visuospatial sketchpad* (Baddeley, 2012). A later addition was a third STM slave system, the *episodic buffer* (Baddeley, 2000).

The phonological loop provides a temporary store for verbal and auditory information, which is maintained for three to four seconds before decay; with the control process of rehearsal preserving information for longer periods (Baddeley, 2012). Visual, spatial and kinaesthetic information are temporarily stored in the visuospatial sketchpad (Baddeley, 2012). Whilst the episodic buffer is assumed to have the capacity to integrate information from both STM stores, long-term memory (LTM), and perception; to form multidimensional coded episodes (Baddeley, 2000). The central executive is the crux of the WM system, retrieving information from LTM, and managing and coordinating the operation of the visuospatial sketchpad, phonological loop and episodic buffer (Baddeley, 2012).

Cowan's embedded process theory (Cowan, 2010b) is an alternative WM model, which in many respects is similar to the multi-component model (Baddeley, 2012). A fundamental difference is that it comprises a LTM based unitary storage system, at various levels of activation or attentional focus (Baddeley, 2012). The phonological and visuospatial stores are considered instances of temporary activated LTM and are equivalent to STM (Cowan, 1993). Also, the activated memory categories are not limited to auditory and visuospatial, but extend to include the senses such as taste and smell, and orthographic and semantic language features (Cowan, 2010b). Dissimilar categories of activated memory can coexist independently, whereas interference occurs between similar categories (Cowan, 2010b). A central executive is responsible for cognitive control and the focus of attention, which is limited to maintaining approximately four items or chunks in a readily accessible or hyper-activated state from the activated memory set (Cowan, 2010b). Information displaced from attentional focus remains temporarily activated (Cowan, 2010b).

More recently, Unsworth and colleagues (Unsworth & Engle, 2007a, 2007b; Unsworth & Spillers, 2010; Unsworth, Spillers, & Brewer, 2010) presented the dual component model of WM that combines the focus of attention from Cowan's embedded process theory, the neuro-computational model activation buffer (Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005), and the episodic buffer hypothesised by Baddeley. The actively maintained primary memory combined with a searchable, cue indexed secondary memory comprise the two components (Unsworth & Engle, 2007a).

Primary memory is equivalent to the activation set under focused attentional control in Cowan's embedded process theory. Capacity is limited to four separate item representations, but may be reduced depending on task demands (Cowan, 2001, 2010a; Garavan, 1998). Displacement to secondary memory occurs through the addition of new incoming item representations or distraction by a secondary task (Unsworth & Engle, 2007a). In either case, items are displaced from primary memory to secondary memory through the removal of attention.

Conditional on memory load, the retrieval of displaced items from secondary memory may be a competitive process at times. Therefore to maximise efficiency, a cue based search process is likely to ensure only relevant items are retrieved, with the selection criteria based on permutations of temporal, contextual, and categorical information (Unsworth et al., 2010).

Potential retrieval problems include proactive interference, encoding deficiencies and output interference (Unsworth & Spillers, 2010).

WM is operationalised through WMC. The definition of WMC is model dependent; but very generally WMC is considered “the amount of information an individual can maintain in a particular task that is designed to measure some aspect/s of WM” (Conway, Getz, Macnamara, & Engel de Abreu, 2011, p. 395).

Complex span tasks, including reading span, operations span, and counting span were the original measures of WMC, and based on Baddeley and Hitches’ (1974) multi-component model of WM. All three of these instruments use a series of trials that interleave a number of processing tasks with items to-be-remembered (Conway et al., 2005). The name of each instrument is indicative of the processing component; reading span uses reading processing tasks; operations span, mathematical; and counting span, numerical. At the end of each trial, the to-be-remembered items are required to be recalled. The span is the number of memory items that can be recalled in correct serial order provided the processing component error threshold was not breached, which invalidates the test.

Recall. A characteristic of simple and complex span tasks is the requirement of the subject to recall to-be-remembered information to the test administrator; either orally or in written form (Conway et al., 2005). Simple and complex span tasks demonstrate active recall, where the subject is aware in advance of testing that they will be required to recall information (Ross & Di Vesta, 1976). Other recall categories include free, serial and cued. Free recall requires the subject to recall a list of usually familiar to-be-remembered items in any order using internally generated cues (Tulving & Patterson, 1968). Serial recall requires the order of the to-be-remembered items to be maintained (Cowan, Saults, Elliott, & Moreno, 2002). Lastly, hints or external cues are used in cued recall, to facilitate the retrieval of information (Tulving & Osler, 1968).

Relationships: *N*-Back Tasks, STM, and WM

Neurophysiology: *n*-back tasks and WM. *N*-back tasks have been considered the *gold standard* in functional neuroimaging of brain areas associated with WM (Kane & Engle, 2002). However, much of the evidence surrounding the relationship between *n*-back tasks and WM has been circumstantial. *N*-back tasks were presumed to require WM because of their assumed cognitive requirements of: monitoring, updating and manipulation of remembered information (Owen, McMillan, Laird, & Bullmore, 2005); storage and manipulation of information (Kane & Engle, 2002); and the update of a dynamic rehearsal set whilst simultaneously responding to a continuous stream of stimuli (Kane, Conway, Miura, & Colflesh, 2007).

The above are all reasonable descriptions of the possible cognitive processes utilised in resolving the *n*-back task, yet in some cases the terminology used is vague and the assumed *n*-back task processes have not been subject to any empirical substantiation (Conway et al., 2005; Kane et al., 2007). The underlying basis for these statements is neurophysiological. Kane and Engle (2002) identified the prefrontal cortex as a significant contributor to WM through brain imaging, single dissociation cases, and primate research. Neural functional magnetic resonance imaging (fMRI) and positron emission tomography of the *n*-back task revealed activation of the dorsolateral prefrontal cortex, and therefore based on proximity, the *n*-back task was assumed to be using WM (Kane & Engle, 2002).

N-back tasks have been associated with the dorsolateral prefrontal, mid-ventrolateral prefrontal, parietal, and anterior cingulate cortices (Owen et al., 2005). In a comprehensive meta-analysis of neuroimaging studies, Owen et al. (2005) categorised *n*-back tasks by stimulus type, verbal or non-verbal; and the type of monitoring required, identity or location. Verbal identity monitoring was associated with increased activation of the left ventrolateral prefrontal cortex, an area associated with inner speech (Mesulam, 2000). In contrast, non-verbal location monitoring was associated with enhanced activation of the right dorsolateral prefrontal, lateral premotor, and parietal cortices, which comprise the spatial attention network (Mesulam, 2000).

Chein, Moore, and Conway (2011) have questioned the legitimacy of much of the WM neuroimaging literature. They have argued that the majority of research has failed to employ tasks that fully engaged WM architecture. *N*-back tasks were reasoned to require recognition rather than engaging cued recall as per the dual component theory of WM. A small number of eclectic studies were cited that have engaged complex span tasks, which they considered a more valid method of activating WM, and better predictors of higher cognitive abilities.

Consistent with complex span instruments, Chein et al. (2011) devised tasks that combined verbal or spatial processing with storage components, which required information to be encoded and maintained whilst concurrently managing a processing task (encoding maintenance and coordination phase); and later retrieval of the to-be-remembered items (recall phase). Neural fMRI was used to capture images of both phases.

Increased activity was found in the lateral prefrontal, anterior cingulate and parietal cortices during the encoding maintenance and coordination phase for both verbal and spatial complex span tasks. Activity was found to increase for same-domain tasks relative to cross-domain tasks indicating greater interference which concurs with behavioural studies (Bayliss, Jarrold, Gunn, & Baddeley, 2003). Furthermore, the recruitment of the prefrontal cortex and anterior cingulate cortices in the cross-domain task was only slightly less than the recruitment for the same-domain task, suggesting that these neural areas are used in the engagement of attentional control and selection mechanisms. Bilateral activation of the anterior prefrontal cortex and the medial temporal lobe during the recall phase implicated a cued search of LTM as predicted by Unsworth and Engle (2007a).

Construct validity: *n*-back, WM and Gf. Chein et al. (2011) claimed that their customised tasks fully engaged WM architecture, however they activated almost identical neural structures as documented by Owen et al. (2005) using recognition based *n*-back tasks. Further, Chein et al. did not make use of a control treatment condition, and similarly 20 of the 24 studies documented by Owen et al. (2005) failed to use a control condition also. Thus, it is extremely difficult to conclude that different neural structures are predisposed to WM processes. Moreover, the granularity of fMRI provides no evidence of the cognitive processes invoked.

Kane, Conway, Miura, and Colflesh (2007) investigated the construct validity between *n*-back, WM, and Gf; using a single *n*-back task with memory loads of two and three back. WM and Gf were operationalised by operations span and Raven's Advanced Progressive Matrices (RAPM; Raven, Raven, & Court, 2003) respectively. Single *n*-back accuracy and operations span were found to be only weakly correlated, however both accounted for independent variance in RAPM, implying that single *n*-back and WM as operationalised by operations span do not represent a single construct.

Study 1 of Jaeggi, Studer-Luethi, et al. (2010) had very similar findings in an almost identical study with memory loads of two to four back. Operations Span was weakly but significantly correlated with single *n*-back accuracy, which is in contrast to Jaeggi, Buschkuehl, Perrig, and Meier (2010), who failed to record a significant correlation with reading span. Also, single *n*-back was more strongly correlated with RAPM than operations span. Regression modelling revealed operations span made a negligible contribution to single *n*-back variance. Synonymous with Kane et al. (2007), operations span and single *n*-back were both found to make unique variance contributions to RAPM, with only that of single *n*-back being significant. The outcome of both studies is summarised in Table 1.

Table 1

Summary of Correlations and Shared Variance between Single N-Back Tasks, Operations Span and RAPM for Kane et al. (2007) and Jaeggi, Studer-Luethi, et al. (2010)

	Kane et al. (2007)	Jaeggi, Studer-Luethi, et al. (2010)
1. Single *n*-back task and operations span correlations	r = -.08 to .22*	r = .21*
2. Correlations with RAPM	Single *n*-back r = -.21* to .42* Operations span r = .33*	Single *n*-back r = .44** Operations span r = .24*
3. Unique variance associated with RAPM	Single *n*-back R^2 = .18* to .24* Operations span $R^{2\,=}$ .07*	Single *n*-back R^2 = .19** Operations span $R^{2\,=}$ .03

Note. RAPM = Raven's Advanced Progressive Matrices.
* p < .05, ** p < .01.

Construct validity between the *n*-back task and WMC was realised with two studies conducted by Shelton and colleagues (2009; 2007). They used a modified single *n*-back task where word lists containing four to six items were visually presented in a random order, and participants were asked to recall one to three back with a typed response. Operations span and digit span correlations with *n*-back accuracy were .38 and .48 respectively, and ranged between .33 and .45 with other measures of WM (Shelton et al., 2009). The authors concluded that the recall version of single *n*-back was a valid measure of WM.

All three studies suggested that the underlying cognitive processes of *n*-back and complex span tasks are distinctly different, as emphasised by the weak correlations between single *n*-back tasks and WM as measured by operations span; and the independent variance shared between RAPM and single *n*-back tasks, and RAPM and operations span. The work of Shelton and colleagues indicated that changing the single *n*-back task to use recall invoked processes common to simple and complex span tasks, which is assumed to be active recall.

However, they failed to test the modified single *n*-back task using recognition, which would have ensured the variance shared between the modified single *n*-back task and operation span was associated with active recall and not the experimental treatment process.

Recollection and recall. Jaeggi, Studer-Luethi, et al. (2010) argued different memory processes were responsible for the low correlations between *n*-back tasks and operations span. *N*-back was assumed to rely on passive recognition, whereas operations span requires active recall.

Unsworth and Engle (2007a) presented a summary on recognition and cued recall, where subjects are provided external cues to aid in retrieval. Evidence suggests there are two mechanisms that control performance: a fast automatic familiarity process, and a slower controlled recollection process (Yonelinas, 2002), which are referred to as familiarity and recollection respectively. Thus for simple recognition tasks, familiarity is activated; however finer distinction may require the use of recollection, to recover the required object from memory.

Application of the dual component model of WM, implies a search and retrieval of secondary memory is only required for recollection. The authors reviewed a number of studies (Bunting, Conway, & Heitz, 2004; Conway & Engle, 1994; Oberauer, 2005) finding that individual differences in WMC were only realised when recollection was invoked. Furthermore, Oberauer (2005) found evidence to suggest that recall and recognition as measured by complex span tasks use either independent WMC variance sources, or are only partially sharing the same source of WMC variance.

Oberauer’s (2005) finding provides a possible explanation for the weak relationship between *n*-back and complex span tasks. Further, the interaction between *n*-back memory load and the recognition processes of familiarity and recollection is an area for future investigation.

Latent variable analysis: STM, WMC. Kane et al. (2004; Unsworth & Engle, 2007b) conducted a latent variable study examining the relationship between visuospatial and verbal, STM and WMC. WMC was found to be a unitary construct across the verbal and spatial domains, with verbal and spatial WMC sharing at least 70% of their variance. WMC capacity and STM were found to be separate yet highly correlated constructs, with a shared variance of 62.4% and 79.2% for the verbal and spatial domains respectively. Verbal and spatial STM was found to share 40% of their variance.

The sharing of construct variance between verbal and spatial domains is important for single *n*-back near transfer measurement in the current study. Single *n*-back is a visuospatial task, however the instruments that will be used to measure WM and STM require recall and therefore invoke verbal processes. Shared variance between verbal and spatial domains implies that it is plausible for a visuospatial training task to influence a verbal measure, such as backward digit span.

Given the considerable shared variance between WMC and STM in both verbal and spatial domains, it is puzzling that dual *n*-back training can realise a significant increase in STM without affecting reading span, a measure of WMC (Jaeggi et al., 2008). Further, the effect size associated with STM performance improvement was large. The current understanding of *n*-back transfer properties is insufficient to explain this outcome, and replication may be the best avenue of approach to ensure it is indeed reliable.

Summary

From a single study, STM was found to be a near transfer mechanism for dual *n*-back training (Jaeggi et al., 2008). However, near transfer to WM as operationalised by complex span tasks has proven inconsistent with either single or dual *n*-back training. Possible explanations include; lack of a standardised *n*-back application; varying *n*-back modes; use of different stimuli types within the *n*-back task; varying *n*-back memory loads; varying training durations; and internal validity concerns.

To resolve the *n*-back task, some form of memory processing is required. Neuroimaging studies of the *n*-back task and WM provide only circumstantial evidence that equivalent neural processes are being activated. The use of complex span tasks also appear to be of limited benefit to advancing the understanding of near transfer mechanisms. Correlational and regression studies have revealed at best a significant but weak relationship between complex span measures and single *n*-back tasks, with regression analysis indicating that the variance shared between WMC and Gf is independent of that shared by single *n*-back tasks and Gf. Furthermore, the modification of the single *n*-back task from recognition to one that requires the active recall of information by Shelton and colleagues (2009; 2007) significantly improved the correlation of the single *n*-back task with measures of STM and WMC; and emphasised that the processes that are used in active recall differ from those that are required for recognition.

Purpose of Present Study

The purpose of the present study was to investigate STM and WM as possible near transfer mechanisms of single *n*-back training. To date, only STM has been significantly improved with dual *n*-back training, with most studies finding no improvement in WM as operationalised by complex span measures. Sternberg (2008) and others have stressed that the relationship between *n*-back tasks and WM required a more diverse investigation. Indeed, it may be that a systematic understanding will only be achieved by studying performance on a battery of tests of WM, and for that matter STM (Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008). The present study was designed to investigate the role of WM and STM as defined by CHC theory and operationalised by tests of the WJ-III (Mather & Woodcock-Johnson, 2001).

Adoption of the CHC model provided a methodical and theoretically based approach to identifying the near transfer mechanisms of the single *n*-back task. STM (or Gsm under CHC theory) is one of the 16 broad second order abilities and is operationalised by backwards digit span (Mather & Woodcock-Johnson, 2001) which is congruent with Jaeggi et al. (2008) and latent variable analysis studies where both forward and backwards span instruments have been used (Conway et al., 2005; Kane et al., 2004; Unsworth & Engle, 2007b). However, WM refers to the cognitive ability to mentally manipulate information held in immediate awareness and is tested accordingly (Mather & Woodcock-Johnson, 2001); and therefore differs from WM as operationalised by complex span tasks (Conway et al., 2005; Martínez et al., 2011). Factor analysis has identified that WM is subordinate to Gsm within CHC theory (Mather & Woodcock-Johnson, 2001).

Given the significant improvement in STM using a dual *n*-back training regime (Jaeggi et al., 2008), there was a cautious expectation that there would be a similar outcome for single *n*-back training. The subordinate relationship of WM to STM also provided optimism that single *n*-back training would improve WM. However, the inconsistency of WM performance improvement from previous *n*-back training studies, and the non-replication of the improvement in STM from dual *n*-back training provided support for using non-directional statistical tests.

The current study mirrored the single *n*-back training regime of Jaeggi, Buschkuehl, Jonides, and Shah (2011) and focused on the constructs of STM and WM. Tests of the WJ-III were used to investigate the effect of single *n*-back training on STM; and to further investigate the effect of single *n*-back training on WM. It was hypothesised that single *n*-back training would result in a significant difference between pre- and post-training STM scores. Moreover, given the subordinate relationship of WM to STM within the CHC theoretical framework it was also hypothesised that single *n*-back training would result in a significant difference between pre- and post-training WM scores.

Method

Design

The current study utilised a two factor mixed design. The between-subjects factor was *training type* and the within-subjects factor was *time of testing*, with each factor containing two levels. Training type comprised single *n*-back and general knowledge cognitive training regimes. Time of testing comprised pre- and post-training.

The experiment was divided into three contiguous phases: pre-training, training and post-training. STM and WM were the two dependent variables measured at pre- and post-training.

Participants

Potential participants were recruited from undergraduate psychology courses at Edith Cowan University (ECU), Western Australia and the personal network of the researcher. Seventy-four people volunteered to participate in the study; however, 26 (15 from the experimental treatment group and 11 from the active control group) had withdrawn or were non-contactable prior to the post-training phase dependent variable testing. Further, one other was removed from the sample as their training data appeared to be fabricated when compared to the other participants. The participants comprised 30 females and 17 males, with the ages of both females and males ranging from 18 to 69 years (M_{female} = 38.78, SD = 12.31; M_{male} = 36.57, SD = 14.89).

Power calculations were used to estimate the required sample size. Transfer to STM identified by Jaeggi et al. (2008) using dual *n*-back training had a Cohen's effect size of f =

0.45 which is considered large. However, the more conservative effect size, *d* = 0.65 associated with single *n*-back transfer to Gf was used (Jaeggi, Studer-Luethi, et al., 2010) as the possible transfer to WM was unknown and was likely to be smaller (Jaeggi et al., 2008; Jaeggi, Studer-Luethi, et al., 2010). Therefore, it was estimated that the experimental treatment and active control groups would each require 25 participants to achieve a statistical power of 0.82 (Faul, Erdfelder, & Lang, 2010, January 1; Field, 2009).

Pre-Training Phase Procedure

Upon ECU Human Research Ethics Committee approval, potential participants were provided with the participant information letter (see Appendix A). After ensuring that the participants were cognisant of the research requirements and subject to signing the informed consent form (see Appendix A), participants were randomly assigned to either the experimental treatment or active control groups.

Following the gathering of demographic information (see Appendix A), Test 7: Numbers Reversed and Test 9: Auditory Working Memory from the WJ-III were administered to the participants. In addition, participants were also administered WJ-III, Test 13: Picture Recognition as part of a companion study researching the effect of cognitive training on visual WM (Prandl, 2012). The reliability coefficients for Tests 7 and 9 range from .88 to .93 for ages 20 through to 79 years (McGrew & Woodcock-Johnson, 2001).

WJ-III, Test 7: Numbers Reversed. Numbers Reversed is an individually administered 30-item test that required the participant to perform the mental operation of reversing the order of a span of numbers held within memory. Item spans varied in length between two and nine digits, and were delivered to the participants in order of ascending span length. Absolute scoring was used to collect responses. Numbers Reversed is primarily a test of STM (Gsm); however, it is also classified as a measure of attentional capacity (Mather & Woodcock-Johnson, 2001).

WJ-III, Test 9: Auditory Working Memory. Auditory Working Memory is a 21-item instrument that required the participant to hold a list of words and numbers within memory, which were then required to be recited in categorical serial order, words first followed by numbers. Item spans varied in length between two and nine objects, and were delivered to the participants in order of ascending span length. Absolute scoring was used to collect responses. Auditory Working Memory is a measure a STM span, and can also be classified as a measure of divided attention (Mather & Woodcock-Johnson, 2001).

Tests 7, 9 and 13 were administered consecutively to each participant. The test order was counter-balanced across the participant sample to mitigate order effects. The process of meeting, explaining the research, and administering the tests to each participant took between 30 minutes and one hour.

WJ-III, Working Memory. Working Memory denotes the ability to cognitively manipulate information that is being held in immediate awareness (Mather & Woodcock-Johnson, 2001). It is an additional clinical cluster which is calculated by obtaining the mean of the *W* scores of Test 7 and Test 9, with a reliability coefficient ranging between .89 to .94 for ages 20 through to 79 years (McGrew & Woodcock-Johnson, 2001). *W* scores are a transformation of the Rasch ability scales which standardises disparate WJ-III Test scores onto an equivalent scale, with a mean of 500 (Mather & Woodcock-Johnson, 2001).

Training Phase Procedure

Subsequent to the pre-training phase, the single *n*-back task was assigned to participants in the experimental treatment group; and the tasks of *Definetime* and *Who Wants to Be a Millionaire* (Millionaire) were assigned to those in the active control group.

Explaining the designated cognitive task, loading the training software onto a participant's personal computer, and monitoring a practice round took approximately 20 minutes for each participant. The following is a description of the cognitive tasks, training duration and scoring practices.

Single *n*-back. The single *n*-back task is an interactive computer based application which was downloaded from the Brain Workshop website (Hoskinson, 2008). The Brain Workshop application is similar but not identical to the software developed by Jaeggi et al. (2010). It uses the same underlying Brain Twister algorithm and is sufficiently customisable to provide an identical graphical user interface and scoring process (see Appendix B for details). Subsequent to customisation, the software was packaged onto a USB flash drive to allow for rapid installation onto participant personal computers (see Appendix C).

Figure 1. Single *3*-back sequence.

The single *n*-back task is a memory activity requiring the participant to remember where a visuospatial stimulus or target was presented *n* iterations previously. The visuospatial stimulus is a blue square that randomly accommodates one of eight positions on a black background for a period of 500ms, followed by a 2500ms interstimulus interval. This 3000ms sequence is defined as a trial. A block consists of 20 + *n* trials, where there are six visual targets per block. A response is required whenever the visuospatial stimulus matches the stimulus *n* iterations back in the sequence. To register a match, the <A> key is pressed. No response is required for non-targets. Figure 1 is an example of a *3*-back sequence.

The single *n*-back software was configured to always start each daily training session at *2*-back. The *n*-back task automatically manages the level of difficulty based on the score from the previous block by altering the memory load, *n*. If the participant made fewer than three mistakes, *n* was increased by one; if five or more mistakes were made *n* was decreased by one; otherwise, the difficulty level remained unchanged. This reflected a performance of greater-than-or-equal to 90% or less-than-or-equal to 70% respectively. Performance is measured using the following formula:

$$\text{Performance}=\frac{(\text{True positives}+\text{True negatives})\times 100}{(\text{True positives}+\text{True negatives}+\text{False positives}+\text{False negatives})}$$

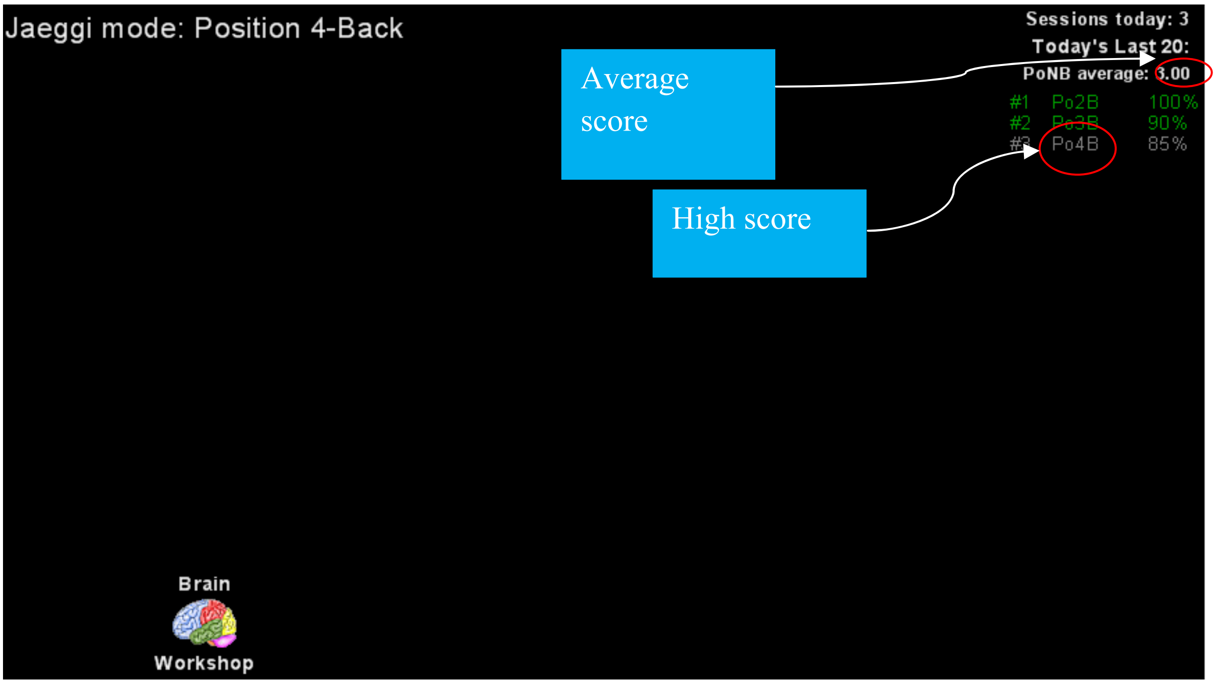

Figure 2. Average and high score locations on the single *n*-back secondary panel.

Each participant was required to complete 20 sessions over a contiguous 30-day period. A training session consisted of 15 rounds of single *n*-back, which took between 15 and 20 minutes to complete. The advised training routine was five consecutive days followed by a two-day break. On the scoring sheet provided, participants recorded their average and highest score (see Figure 2) attained in each session.

Definetime. Definetime (East of the Web, 2011) is an internet based quiz that requires the participant to select the correct definition of a word from four possible options. A correct response gives the player 10 points, however an incorrect response reduces the player's score by five points. Players must continue to choose a response until they select the correct definition. The object is to score as many points as possible within the two-minute time limit.

Each participant was required to complete 20 sessions over a contiguous 30-day period. A training session was 20 minutes in length alternating between Definetime and Millionaire tasks. Definetime automatically scores each game (see Figure 3). Participants were asked to record the highest score for each session on the scoring sheet provided.

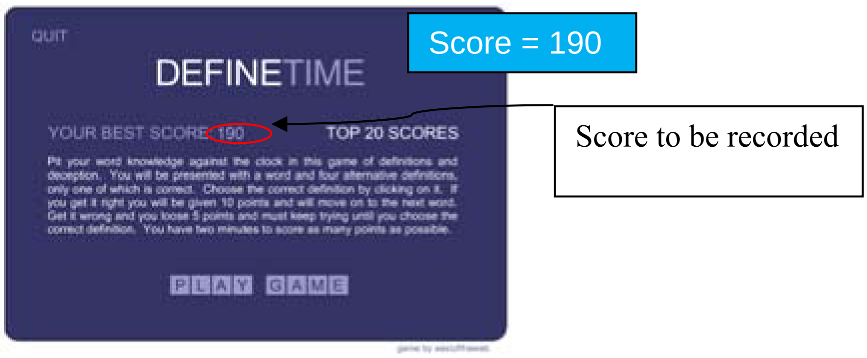

Figure 3. Score location on the Definetime primary panel.

Millionaire Who wants to be a millionaire (Box10, n.d.) is an internet version of Channel 9's game show hosted by Eddy McGuire. The objective is to win one million dollars by answering each of the 15 consecutive questions correctly within the 30-second timeframe. If a participant is unsure of the answer, they have three lifelines per game that can be chosen in any order to improve their probability of answering correctly. The first lifeline is *50% chance*, where two of the potential answers are removed. The second lifeline is *call-a-buddy*, where the participant can refer the question to a fictional friend. They provide a possible answer along with the probability that their answer is correct. Finally, the third lifeline is *audience assistance*, where the participant can ask the audience for help.

Each participant was required to complete 20 sessions over a contiguous 30-day period. A training session was 20 minutes in length, alternating between Definetime and Millionaire tasks. For Millionaire, every correct answer within a game earns one point. There is a maximum of 15 points per game. Participants were required to record the highest score for each session (see Figure 4) on the scoring sheet provided.

Figure 4. Score location on the Millionaire primary panel.

Post-Training Phase Procedure

A meeting was arranged with each participant within three days of the culmination of the 30-day training period where practicable. Completed participant scoring sheets were collected, and participants were re-tested using Test 7: Numbers Reversed and Test 9: Auditory Working Memory. Written feedback in the form of percentile rankings for WM and visual memory (Prandl, 2012) was provided for each participant's pre-training phase test results (see Appendix A). Following completion of all participant testing, data was collated for analysis and processed using PASW Statistics, Version 20.

Analysis

General. Table 2 summarises the statistical tests used in the current research and their required assumptions. Normality was tested using the Shapiro-Wilk statistic, and visually using detrended and normal Q-Q plots; homogeneity of variance with Levene's test of homogeneity of variances, or F_{max}; homogeneity of covariances with Box's test of equality of covariance; homogeneity of regression slopes with analysis of variance (ANOVA); and linearity with the visual inspection of the graphical depiction of covariate and dependent variable relationships. Only statistical test assumption violations were reported in the Results section.

Table 2

Inferential Statistical Tests Used and their Required Assumptions

Statistical Test	Assumptions
All	Normality
ANCOVA	Homogeneity of regression slopes Linearity Homogeneity of variance
ANOVA	Homogeneity of variance
Independent samples *t*-test	Homogeneity of variance
Paired samples *t*-test	Normality of difference scores Homogeneity of variance
SPANOVA	Homogeneity of variance Homogeneity of covariance

Note. ANCOVA = Analysis of covariance, ANOVA = Analysis of variance, SPANOVA = Split-plot analysis of variance.

Influence of WJ-III test order. The administration of WJ-III tests to participants was counter-balanced in an attempt to mitigate any order effects. Identification of possible order effects for Test 7 and Test 9 (Test 13 was also administered for a companion research project) at both the pre- and post-training phases from the six possible test order combinations was achieved using ANOVA. The Shapiro-Wilk test of normality was violated for Test 7 and test order 7/13/9 at the pre-training phase, with $W(6) = .700$, $p = .006$, indicating a non-normal distribution, however ANOVA is considered robust to violations of normality. Comparisons of test order for Test 7 and Test 9 for the pre-training phase, $F(5, 41) = 0.545$, $p = .741$ and $F(5, 41) = 1.204$, $p = .325$; and post-training phase, $F(5, 41) = 0.659$, $p = .166$ and $F(5, 41) = 0.913$, $p = .482$ respectively were found to be non-significant. In addition a non-parametric Kruskal-Wallis test was conducted for pre-training Test 7, which also produced a non-significant result, $H = 6.170$ (5, $N = 47$), $p = .290$. The non-significant results implied that the test results were independent of order effects.

Exploration of data by group treatment condition. WJ-III Test 7: Numbers Reversed, Test 9: Auditory Working Memory, and the WM composite cluster measures from the pre- and post-training phases were examined by group treatment for normality and the presence of outliers. All distributions were found to be normal, however outliers were identified in pre-training Test 9 (participants B07 and B17), and in post-training Test 9 (participant B14), and pre-training WM (participant B17) for the active control group. Due to the small sample sizes, only outliers that were extreme (i.e. exceeded three box lengths on a box and whisker plot) were further analysed for their influence on statistics.

Results

Descriptive Statistics for WJ-III Tests 7, 9 and WM Scores

Table 3 displays the descriptive statistics for the WJ-III Tests 7, 9 and WM scores at the pre- and post-training phases. Visual inspection of the means indicated that the active control group out-performed the experimental treatment group at the pre-training phase. Allocation of participants to treatment groups was random, so this occurrence was purely coincidental. A comparison of treatment group mean scores for each of the WJ-III tests at the pre-training phase using an independent samples *t*-test (see Table 4) were found to be non-significant; indicating there was no difference between experimental treatment and active control group pre-training phase means.

Table 3

Pre- and Post-Training Phase Descriptive Statistics for the WJ-III Tests 7, 9 and WM Scores

		Pre-training		Post-training	
Treatment group	*n*	*M*	*SD*	*M*	*SD*
		Test 7			
Experimental	26	17.15	3.94	18.15	4.29
Control	21	17.48	3.90	18.43	4.00
Total	47	17.30	3.55	18.28	4.12
		Test 9			
Experimental	26	28.73	5.45	32.42	5.76
Control	21	30.71	4.72	32.71	5.81
Total	47	29.62	5.18	32.55	5.72
		WM			
Experimental	26	530.58	15.29	538.19	17.71
Control	21	534.29	12.68	539.43	16.39
Total	47	532.23	14.16	538.74	16.96

Table 4

Comparison of Pre-Training Phase Experimental Treatment and Active Control Group Mean Scores for WJ-III Tests 7, 9 and WM

	Independent samples *t*-test			
Test	*t*	*df*	*p*	*d*
Test 7	0.306	45	.761	0.092
Test 9	1.316	45	.195	0.385
WM	0.891	45	.378	0.261

WJ-III Tests 7, 9 and WM Performance Changes between Pre- and Post-Training

Performance changes in WJ-III Tests 7, 9 and WM as measured at the pre- and post-training phases for the experimental treatment and active control groups were assessed using a two way mixed split-plot design analysis of variance (SPANOVA), and a one-way analysis of covariance (ANCOVA). In addition, post-hoc paired samples *t*-tests were used to calculate Test performance for experimental treatment and active control groups.

Table 5

Split Plot ANOVA of Pre- and Post-Training WJ-III Tests 7, 9 and WM Scores for the Experimental Treatment and Active Control Groups

				SPANOVA			
Test	Shapiro-Wilk	Box's Test (*p*)	F_{max}	*F*	*df*	*p*	η^2
Test 7	NS	.427	1.922	0.003	1,45	.957	< .001
Test 9	NS	.787	1.511	1.511	1,45	.126	.051
WM	NS	.549	1.949	0.687	1,45	.412	.015

Note. NS = Non-significant.

Test 7: Numbers Reversed. The interaction between pre- and post-training phase Test 7 scores and group treatment was not significant (see Table 5). This implied that both the single *n*-back and active control training tasks led to similar performance changes. The performance improvement for the experimental treatment group, $t(25) = 1.944$, $p = .063$, $d = 0.24$ and the active control group, $t(20) = 1.284$, $p = .214$, $d = 0.27$, was found to be non-significant with both groups realising a small effect size.

Data for Test 7 failed to satisfy the requirements for ANCOVA. The assumption of homogeneity of regression slopes was met, $F(1, 43) = 0.283$, $p = .597$, however the assumption of linearity was violated.

Test 9: Auditory Working Memory. The interaction between pre- and post-training phase Test 9 scores and experimental treatment and active control groups was non-significant (see Table 5). This outcome signals that the experimental treatment and active control groups realised similar performance changes for Test 9. The performance improvement for the experimental treatment, $t(25) = 5.476$, $p < .001$, $d = 0.66$ and active control groups, $t(20) = 2.291$, $p = .033$, $d = 0.38$ were both significant, realising medium and small-medium effect sizes respectively.

ANCOVA processing found pre-training phase scores to be significantly related to post-training phase scores, $F(1, 44) = 66.830$, $p < .001$. However, after partialing out the pre-training phase scores, comparison of the experimental treatment and active control group post-training phase scores for Test 9, $F(1, 44) = 1.714$, $p = .197$, $\eta^2 = .038$ was not significant. This result further supported the previous finding that the experimental treatment and active control groups realised similar performance changes for Test 9.

WM. The interaction between pre- and post-training phase WM scores and experimental treatment and active control groups was not significant (see Table 5). This implies WM changed similarly for the experimental treatment and active control groups between pre- and post-training. The performance improvement for the experimental treatment group was found to be significant, $t(25) = 4.333$, $p < .001$, $d = 0.46$; whereas the performance improvement for the active control group was non-significant, $t(20) = 2.050$, $p = .054$, $d = 0.35$. The experimental treatment group realised a medium effect size, whereas the effect size for the active control group was small-to-medium.

ANCOVA processing found pre-training phase scores to be significantly related to post-training phase scores, $F(1, 44) = 81.139$, $p < .001$. However after partialing out the pre-training phase scores, comparison of the experimental treatment and active control group post-training phase scores for WM, $F(1, 44) = 0.608$, $p = .440$, $\eta^2 = .014$ was not significant. This outcome further supported the previous finding that there was no difference in WM performance between the experimental treatment and active control groups.

Individual Differences in Training Performance and WJ-III Tests 7, 9 and WM Transfer

Jaeggi et al. (2011) examined the individual differences in single *n*-back training performance and transfer to Gf. Treatment groups were separated into small and large training gain subgroups. Significant transfer to Gf was only found for the large training gain single *n*-back subgroup. The WJ-III Tests were evaluated similarly. Individual differences in training performance were established by ranking the difference of the means of the first and last two training sessions, and splitting the ranked differences at the median to create small and large training gain subgroups. The active control subgroups were based on the Definetime task, as this had the larger training performance improvement effect size of the two active control tasks.

WJ-III Tests 7, 9 and WM performance changes for the experimental treatment and active control groups' small and large training gain subgroups, were each assessed using a SPANOVA. Test 7 violated the assumption of normality for the Definetime large training gain subgroup with a non-normal Shapiro-Wilk statistic, $W(10) = .796$, $p = .013$. Visual inspection revealed the distribution was reasonably normal, and it was assumed the SPANOVA statistic was sufficiently robust to manage the violation.

The interactions for Tests 7, 9 and WM; and the experimental treatment and active control training gain subgroups was found to be non-significant (see Table 6). This implies performance in Tests 7, 9 and WM was independent of individual differences in training performance, within the experimental treatment and active control groups.

Table 6

Split Plot ANOVA of Pre and Post-Training WJ-III Tests 7, 9 and WM Scores for the Small and Large Training Gain Experimental Treatment and Active Control Subgroups

				SPANOVA			
Test	Shapiro-Wilk	Box's Test (*p*)	F_{max}	*F*	*df*	*p*	η^2
Test 7	S	.085	2.548	1.011	3,43	.397	.066
Test 9	NS	.179	2.389	1.308	3,43	.284	.084
WM	NS	.569	2.310	0.569	3,43	.625	.040

Note. NS = Non-significant, S = Significant.

Training Phase Performance Changes

Training phase performance for the experimental treatment and active control groups was assessed using a two-tailed paired samples *t*-test to compare the means of the first and last two training sessions for each participant.

Experimental treatment group. The mean *n*-back training performance between the first ($M_{1,2}$ = 3.22, *SD* = 0.70) and last two ($M_{19,20}$ = 4.81, *SD* = 1.39) training sessions was found to improve significantly, $t(25) = 8.006$, $p < .001$ with a large effect size, $d = 1.21$. Training performance improved steadily over the 20-sessions, as displayed in Figure 5.

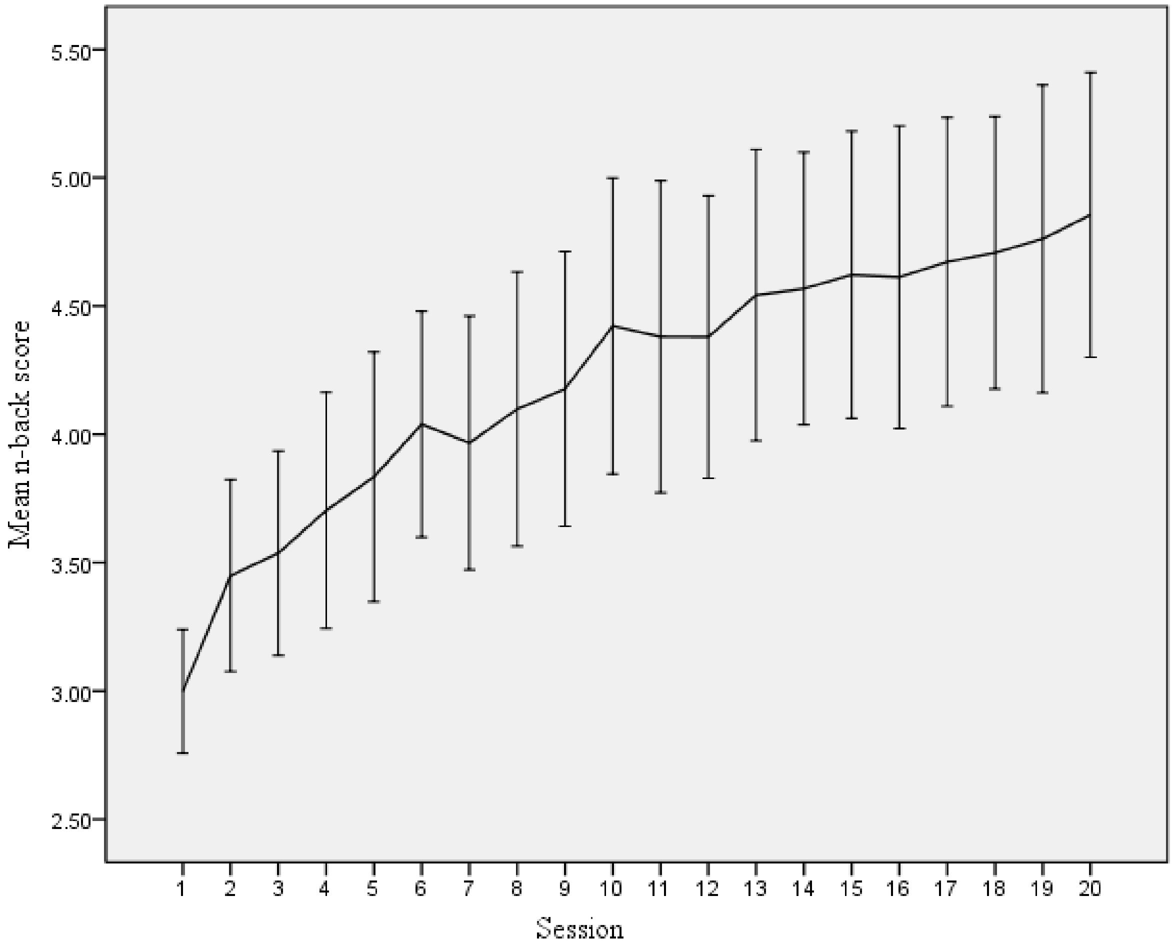

Figure 5. Mean single *n*-back training performance over the 20-session training phase. Error bars reflect the 95% confidence interval.

Active control group. Similarly, there was an improvement in mean training performance for Definetime ($M_{1,2}$ = 236.07, *SD* = 105.41; $M_{19,20}$ = 486.31, *SD* = 211.31) and Millionaire ($M_{1,2}$ = 7.74, *SD* = 1.93; $M_{19,20}$ = 9.14, *SD* = 1.70) between the first and last two training sessions. Training performance improvement for Definetime, $t(20) = 7.473$, $p < .001$

and Millionaire, $t(20) = 2.904$, $p = .009$ was found to be significant, with respective large effect sizes, $d = 1.58$ and $d = 0.77$. Definetime training produced a steady improvement in performance, whereas Millionaire training performance improvement was non-monotonic over the 20-sessions as displayed in Figure 6 and Figure 7 respectively.

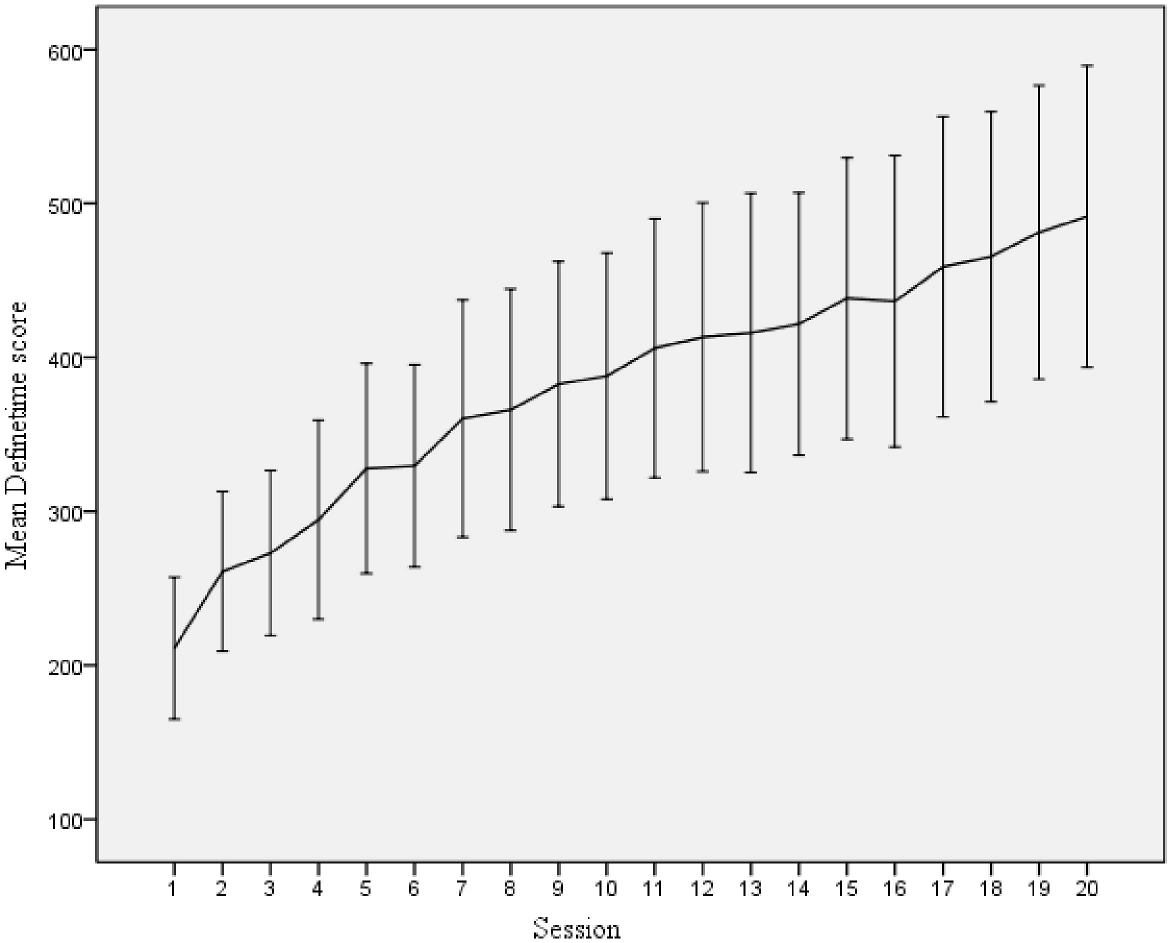

Figure 6. Mean Definetime training session performance over the 20-session training period. Error bars reflect the 95% confidence interval.

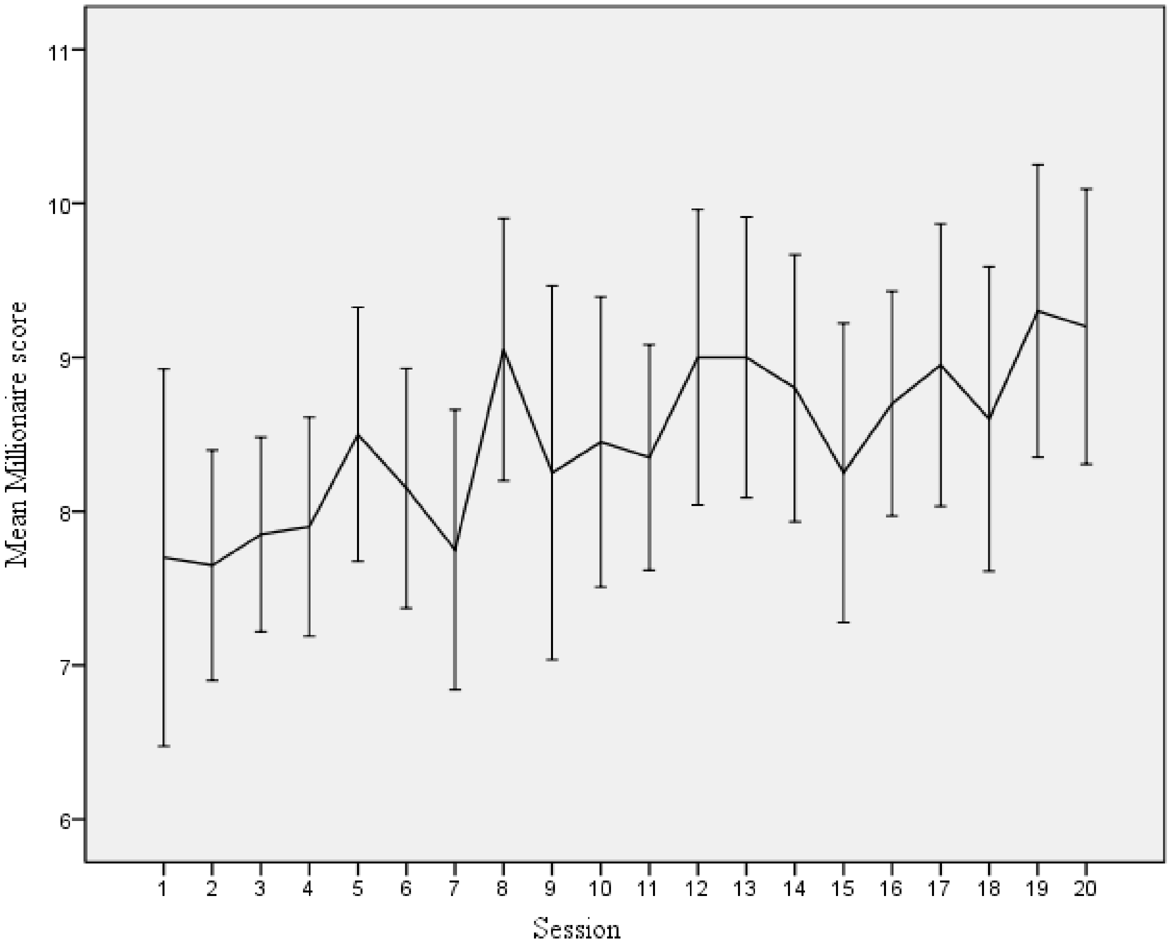

Figure 7. Mean Millionaire training performance over the 20-session training phase. Error bars reflect the 95% confidence interval.

Small and Large Training Gain Subgroup Performance Comparisons

Performance improvements in the small and large training gain subgroups for the experimental treatment and active control groups was analysed using a SPANOVA. The active control group was analysed with respect to both Definetime and Millionaire performance. The results are summarised in Table 7.

Analysis of Table 7 indicates that the effect size of the interaction between small and large training gain performance for the single *n*-back, Definetime and Millionaire tasks was similar. Further, the correlation between participant small and large training gain membership for Definetime and Millionaire was found to be non-significant, $r(21) = -.236$, $p = .302$

implying that being assigned to a subgroup in one active control group task had minimal influence on being assigned to an equivalent subgroup in the other active control group task.

Table 7

Split Plot ANOVA of Small and Large Training Gain Subgroup Training Performance for each of the Three Training Tasks

				SPANOVA			
Treatment	Shapiro-Wilk	Box's Test (p)	F_{max}	F	df	p	η^2
Single *n*-back	NS	.407	1.622	54.117	1, 24	< .001	.693
Definetime	S	.080	4.800	29.223	1, 19	< .001	.606
Millionaire	S	.102	7.401	32.703	1, 19	< .001	.633

Note. NS = Non-significant, S = Significant.

Single *n*-back comparison. The interaction between the small and large training gain subgroups and training performance was found to be significant for the single *n*-back task (see Table 7). Investigation of the means revealed that the large training gain subgroup ($M_{1,2}$ = 3.40, SD = 0.73; $M_{19,20}$ = 5.81, SD = 1.01) significantly outperformed the small training gain subgroup ($M_{1,2}$ = 3.05, SD = 0.64; $M_{19,20}$ = 3.81, SD = 0.93). Post-hoc testing found training performance improvement for the large, $t(12) = 13.094$, $p < .001$, $d = 2.80$ and small, $t(12) = 5.955$, $p < .001$, $d = 0.97$ training gain subgroups was significant. Both subgroups realised large effect sizes. This finding is contrary to Jaeggi et al. (2011) who found only the large training gain subgroup improved significantly for the single *n*-back task. Mean single *n*-back subgroup performance over the 20-training sessions is displayed in Figure 8.

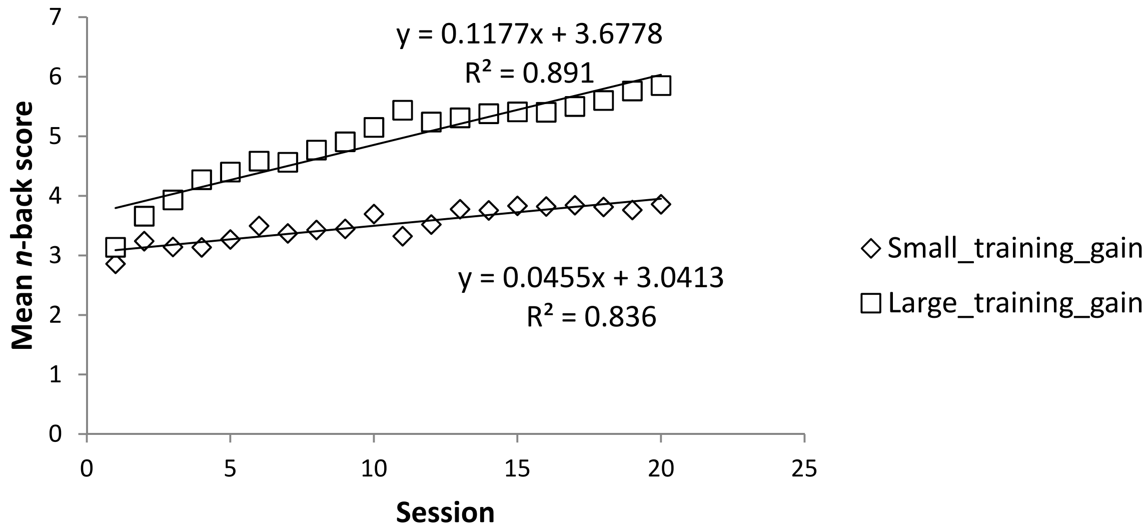

Figure 8. Mean single *n*-back small and large training gain subgroup scores over the 20-session training phase. Large coefficients of determination values indicate a significant amount of performance variance was accounted for by training session.

Definetime comparison. Participant B31 was considered an extreme outlier for the means of the first two sessions' scores in the small training gain subgroup. In addition, distribution of the means of the last two sessions' scores for the large training gain subgroup was not normal, $W(10) = .832$, $p = .036$. Visual inspection of the data indicates neither the outlier nor the non-normal distribution should affect the SPANOVA statistic.

The interaction between the small and large training gain subgroups and training performance was found to be significant (see Table 7). Removal of the outlier brought about a similar result, $F(1, 18) = 26.052$, $p < .001$, $\eta^2 = .591$. Investigation of the means revealed the large training gain group ($M_{1,2} = 298.50$, $SD = 114.26$; $M_{19,20} = 671.00$, $SD = 103.44$) significantly outperformed the small training gain group ($M_{1,2} = 179.32$, $SD = 55.07$; $M_{19,20} = 318.41$, $SD = 120.65$). Both the large, $t(9) = 11.20$, $p < .001$, $d = 6.33$ and small, $t(10) = 4.975$, $p = .001$, $d = 1.58$ training gain subgroups improved significantly. Both subgroups realised large effect sizes. Mean Definetime subgroup performance over the 20 sessions is displayed in Figure 9.

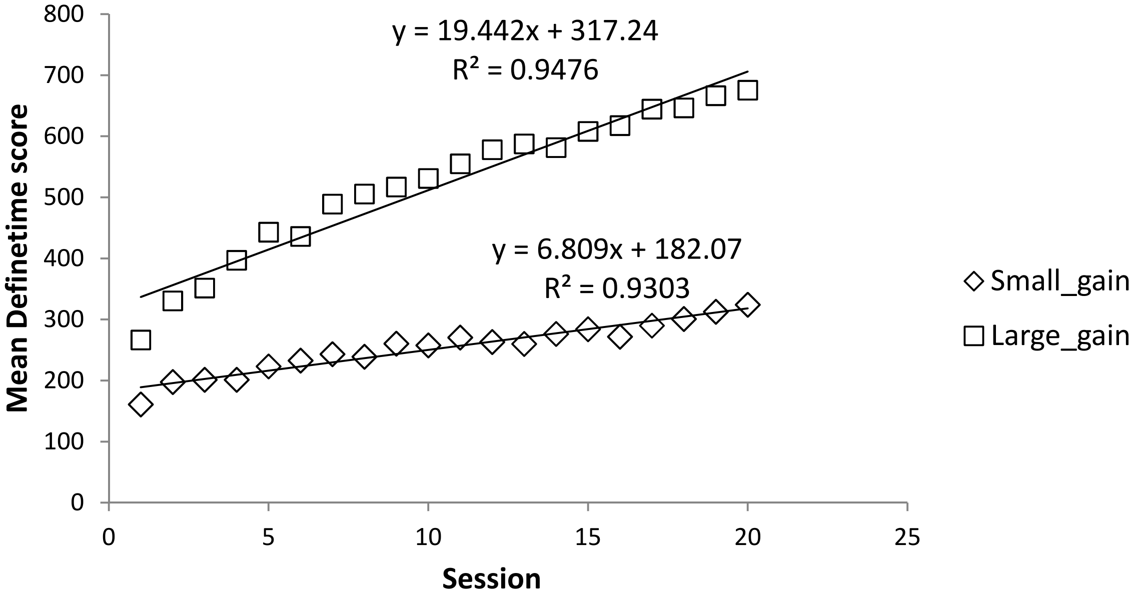

Figure 9. Mean Definetime small and large training gain subgroup scores over the 20-session training phase. Large coefficients of determination values indicate a significant amount of performance variance was accounted for by training session.

Millionaire comparison. Exploration of data revealed B07 was an extreme outlier for the small training gain subgroup, however it was maintained within the sample as removal caused Box's test to become significant (p = .008). Assumptions of homogeneity of variance and co-variance were supported, however the assumption of normality failed for the means of the first two sessions for the small training gain group, $W(9) = .747$, $p = .005$. It was assumed that SPANOVA was sufficiently robust to manage the violation of normality.

The interaction between the small and large training gain subgroups and training performance was found to be significant (see Table 7). Investigation of the means revealed the large training gain subgroup ($M_{1,2} = 6.77$, $SD = 1.85$; $M_{19,20} = 9.82$, $SD = 2.03$) significantly outperformed the small training gain subgroup ($M_{1,2} = 8.80$, $SD = 1.44$; $M_{19,20} = 8.40$, $SD = 0.81$). The improvement in the high training gain subgroup $t(10) = 7.381$, $p < .001$, $d = 1.57$ was found to be significant with a large effect size; however improvement in the small training gain subgroup, $t(9) = 0.910$, $p = .387$, $d = 0.36$ was non-significant with a small effect size. Mean Millionaire subgroup performance over the 20 sessions is displayed in Figure 10.

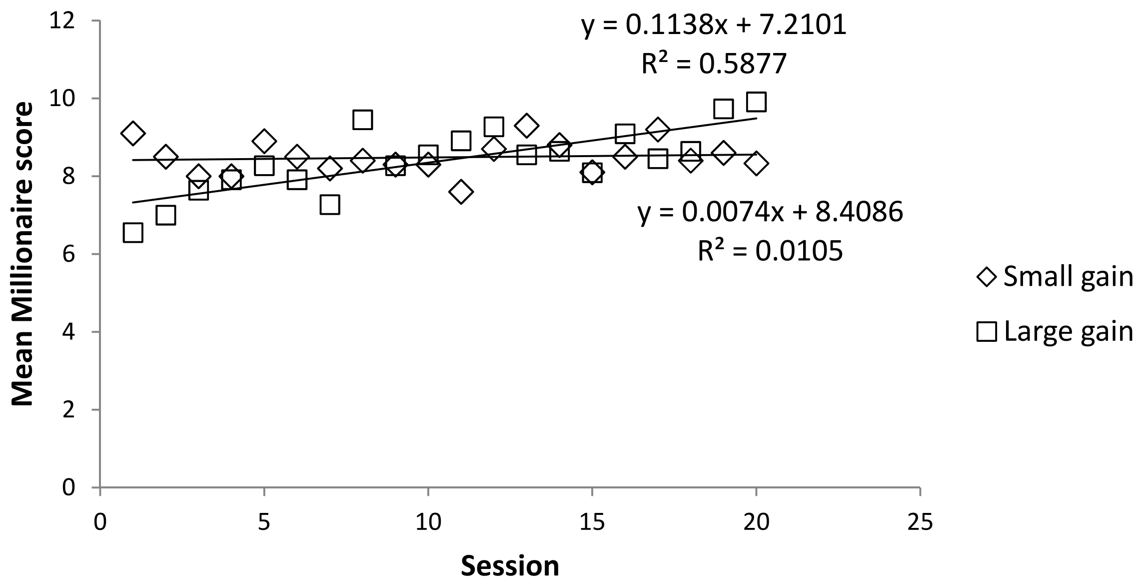

Figure 10. Mean Millionaire small and large training gain subgroup scores over the 20-session training phase. Coefficients of determination values indicate a significant amount of performance variance was accounted for by training session for the large training gain subgroup only.

Discussion

It was hypothesised that single *n*-back training would result in a significant difference in pre- and post-training STM and WM scores relative to the active control tasks, as operationalised by the WJ-III. The resultant difference was non-significant for both constructs, indicating that there was no measurable transfer to STM and WM after 20 sessions of single *n*-back training. Furthermore, after separating treatment conditions into subgroups based on individual differences in single *n*-back and Definetime training performance, comparison of the pre- and post-training STM and WM scores for the subgroups across the two treatment conditions was also found to be non-significant. This finding provides further support for STM and WM as unlikely near transfer mechanisms for single *n*-back training.

The STM result contrasts with Jaeggi et al. (2008), who found a significant improvement in STM with a large effect size. It was assumed that the underlying single *n*-back application algorithm used in the current study was functionally equivalent to that used by Jaeggi and colleagues (2008; 2011; 2010). Experimental control group type and *n*-back modality are the two points of difference between the current research and that of Jaeggi et al.. The current study used an active control group and the single *n*-back task, whereas Jaeggi et al. used a passive control group and the dual *n*-back task. Post-hoc testing of the current research revealed the single *n*-back task obtained a non-significant improvement with a small effect size, with the active control treatment realising a non-significant gain of similar magnitude. Therefore, it is reasonable to conclude that the control group type did not contribute to this outcome and furthermore, it is unlikely STM is a near transfer mechanism of single *n*-back training. Moreover, this finding implicates the auditory component of the dual *n*-back task, either unaccompanied or together with the spatial component to be responsible for improving STM, as measured by forward and backward digit span.

The modification of STM through dual *n*-back training provides a potential near transfer mechanism for far transfer to Gf (Engle et al., 1999; Kane et al., 2004; Unsworth & Engle, 2007b). Further, this implies that near *n*-back transfer mechanisms are modality dependent, and substituting the single *n*-back task for the dual *n*-back task because of its apparent reduced complexity (Jaeggi et al., 2011; Jaeggi, Studer-Luethi, et al., 2010) may not be an equivalent exchange.

The WJ-III definition of WM reflects the cognitive processes necessary to resolve the *n*-back task as described by Owen et al. (2005), Kane and Engle (2002), and Kane et al.

(2007). Therefore, from a face validity perspective, WM would be expected to improve with single *n*-back training. Post-hoc WM testing saw the experimental treatment realise a significant gain with a medium effect size, whereas the active control group realised a non-significant gain of small to medium effect size. There is evidence to suggest this outcome is an artefact of the experimental process.

The current study is identical to Jaeggi et al. (2011) in terms of procedure, differing only in dependent variables measured and the active control task used. Jaeggi et al. was concerned with far transfer to Gf, whereas the current study investigated near transfer to STM and WM. Comparison of the training performance results however demonstrates substantial differences between the two studies. Both studies realised large effect sizes in single *n*-back training performance improvement. However, for the active control group, Jaeggi et al. realised a negligible improvement, whereas for the current study, the effect size was large for both active control group tasks and extremely large for the Definetime task in particular. These differences were further highlighted when investigating individual differences in training task performance. Jaeggi et al. obtained a large effect size for the large training gain subgroup and a medium effect size for the small training gain subgroup for the single *n*-back training task. In comparison, the current study realised large effect sizes for both large and small training gain subgroups. The real point of difference occurred with the active control task where subgroup creation was based on Definetime performance. The effect size for the active control task large training gain subgroup exceeded all other training tasks by more than two fold. The small training gain subgroup also realised a large effect size.

Definetime and Millionaire were chosen as active control tasks in two previous single *n*-back studies (Palmer, 2011; Preece, 2011). Both tasks are web based and were believed to require crystalline knowledge, and therefore were assumed ideal foils for the presumed Gf orientated single *n*-back task. Definetime is a time-limited activity requiring participants to match words with their corresponding definitions; however, the words and definitions are repeated in each training round, and therefore the task shifts from one of knowledge to one of recognition with progressive engagement.

Jaeggi, Studer-Luethi, et al. (2010) have suggested the *n*-back tasks are also based on recognition after finding only weak correlations between the single *n*-back task and operations span; and that the shared variance between operations span and RAPM was independent of the shared variance between the single *n*-back task and RAPM. These findings are a replication of an almost identical study conducted by Kane et al. (2007). Further, if the single *n*-back task is modified to require recall, the relationship with complex span tasks such as operations span becomes significant, sharing some 16% of variance (Shelton et al., 2009; Shelton et al., 2007). Moreover, Oberauer (2005) found evidence to suggest that WMC variance associated with recognition is at least partially independent of WMC variance associated with recall.

Prandl (2012) has identified visual recognition as one of the improved cognitive faculties that is associated with training on Definetime and Millionaire. Thus, the use of Definetime as an active control task may have confounded the findings of the current study. This appears to have affected WM more than STM, as significant transfer to STM would be unlikely even with a passive control group. Whether this is the same type of recognition modified by single *n*-back training is unknown, however there are sufficient grounds for concern.

Limitations

Unsupervised training. Throughout the training phase, participants worked through their cognitive training tasks unsupervised. Ideally, physical supervision of participants would have ensured adherence to the required training structure and processes. Logistically however, it was not possible to have physically supervised training; the required coordination of people was simply unfeasible. Further, monitoring participants through software was not possible either, as there were no documented interfaces that would meet this requirement. The training software that was copied to participants' personal computers for the experimental treatment group, or accessed via a web browser for the active control group, was not owned by the university; and therefore could not be modified for monitoring purposes.

However, the majority of the participants that completed the training had a genuine curiosity toward neural plasticity, and were familiar with the work of Doidge (2010). They were most interested to find out what memory changes had taken place through the cognitive training process. Furthermore, the active control task Definetime appeared to bring out a degree of competitiveness in some of the participants as the best scores each day were posted on the host website. Therefore, the intrinsic motivation generated through prior interest in neural plasticity, and the opportunity to have memory change measured, was likely to cause adherence to the training process as it met their internal goals of discovering more about intelligence and memory (Carr & Dweck, 2011).

These observations are statistically supported by the large performance gains achieved in the experimental treatment and active control group tasks over the 20 training sessions. Single *n*-back and Definetime both attained large effect sizes, whilst Millionaire obtained a medium effect size. Further, the trajectory of the single *n*-back performance change over the 20-sessions was consistent to that of Jaeggi, Studer-Luethi, et al. (2010).

No Passive control group. Jaeggi et al. (2008) found significant dual *n*-back training transfer to STM, but not to reading span. Similarly Jaeggi, Studer-Luethi, et al. (2010) found no transfer to operations span, with either single or dual *n*-back training. Both studies used a passive control group. Sternberg (2008) and others have criticised Jaeggi and colleagues for following this experimental design process as it potentially reduced the internal validity through exposing the studies to extraneous effects such as the placebo effect, Hawthorne effect and expectancy bias. The placebo effect dictates that the that belief in the efficacy of a treatment is sufficient to realise improvement (Kienle & Kiene, 1997). The Hawthorne effect is a by-product of being involved in research, where changes in behaviour and outcomes is independent of the research treatment and dissipates once the experiment has concluded (Sedgwick, 2012). Finally, expectancy bias is the transfer of the researcher's cognitive beliefs about the experiment causing them to unconsciously influence the subjects (Rosenthal & Rubin, 1978). Each of these effects can affect experimental outcomes, yet are independent of treatment. The use of an active control where participants perform a non-related task with an identical set of operating parameters endeavours to mitigate these effects. Any experimental extraneous effects are mirrored in the active control task and are subsequently removed through statistical comparison between the experimental treatment and active control groups.

The current research included an active control group to moderate these effects; however, there were insufficient numbers for the inclusion of a passive control group. In conjunction with the active control group, the passive control allows for the estimation of extraneous effects. It would also have provided an estimation of test-retest reliability, a statistic that is not available for Tests 7, 9 and WM of the WJ-III. This would have assisted in the interpretation of the outcome for STM. The small effect size attained by both experimental treatment and active control groups could be a result of genuine performance improvement, or the consequence of practice effects from pre-training phase testing.

Future Directions

The current research requires replication with Definetime removed from the active control tasks, and included as a separate experimental treatment. This would help substantiate whether the single *n*-back task is indeed transferring to WM as operationalised by the WJ-III. The limited training performance improvement using Millionaire provides support for it being retained as an active control task. In addition, the inclusion of a passive control group would facilitate the understanding of any transfer outcomes.

N-back modality appears to influence near transfer properties to STM, with the current study finding no transfer to STM using single *n*-back training, whereas Jaeggi et al. (2008) reported transfer to STM using dual *n*-back training. This finding requires replication to ensure that this is indeed the case. Further, any future *n*-back research should proceed with the single *n*-back task as it appears to be more elementary in its transfer properties, and therefore is less likely to confound outcomes related to more complex constructs such as Gf and WM.

Finally, the development of *n*-back as free and open-source software would create greater certainty that study findings were not software application dependent. Currently most research indicates *n*-back applications have been developed in-house (Jaeggi et al., 2008; Jaeggi et al., 2011; Jaeggi, Studer-Luethi, et al., 2010; Li et al., 2008), and are therefore likely to manifest idiosyncratic behaviour. Further, this would allow for the specification and testing of standardised settings such as proactive interference and inter-stimulus-interval.

Conclusion

The results have demonstrated that short-term memory and WM as operationalised by the WJ-III are unlikely single *n*-back training near transfer candidates. A non-significant result was found for both constructs after 20-sessions of single *n*-back training over a 30-day period.

Jaeggi, Studer-Luethi, et al. (2010) theorised that the single *n*-back task is recognition based rather than recall. Evidence includes minimal correlations between the single *n*-back and complex span tasks; and independently shared variance between RAPM and the single *n*-back task, and RAPM and complex span tasks (Jaeggi, Studer-Luethi, et al., 2010; Kane et al., 2007). In addition, the modification of the single *n*-back task to one of active recall realises significant shared variance with operation span (Shelton et al., 2009; Shelton et al., 2007), suggesting recognition and recall invoke different cognitive processes.

The inclusion of Definetime in the active control group may have confounded the results. This particularly applies for the near transfer to WM, as single *n*-back training realised a medium effect size, relative to the small to medium effect size for the active control group. At face value, Definetime is a knowledge-based task; however the repetition of word stimuli and possible matching definitions appears to transform the task into one of recognition as more rounds of the task are completed. Therefore, the active control group may have realised a WM training effect that partially negated some of the performance improvement from single *n*-back training.

The non-significant transfer from single *n*-back to STM and the associated small effect size implies this finding was unaffected by the active control group tasks. This result was in direct contrast to Jaeggi (2008), who found dual *n*-back training significantly improved STM. The current finding highlights that for STM at least, *n*-back training transfer is modality dependent. That is dual *n*-back training appears to be a more effective transfer method than single *n*-back training for STM, and/or single and dual *n*-back training are transferring to different cognitive processes.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working Memory and Intelligence: The Same or Different Constructs? *Psychological Bulletin, 131*(1), 30-60. doi:10.1037/0033-2909.131.1.30

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89–195). New York: Academic Press.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1-29. doi: 10.1146/annurev-psych-120710-100422

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 48-79). New York: Academic Press.

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General, 132*(1), 71-92. doi:10.1037/0096-3445.132.1.71

Box10. (n.d.). Who wants to be a millionaire Retrieved February 20, 2012, from http:/www.box10.com/who-wants-to-be-a-millionaire

Bunting, M. F., Conway, A. R. A., & Heitz, R. P. (2004). Individual differences in the fan effect and working memory capacity. *Journal of Memory and Language, 51*(4), 604-622. doi:10.1016/j.jml.2004.07.007

Carr, P. B., & Dweck, C. S. (2011). Intelligence and motivation. In S. R. J. & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 748-770). Cambridge, United Kingdom: Cambridge University Press.

Chein, J. M., Moore, A. B., & Conway, A. R. A. (2011). Domain-general mechanisms of complex working memory span. *NeuroImage, 54*(1), 550-559. doi:10.1016/j.neuroimage.2010.07.067

Colom, R., Abad, F. J., Quiroga, M. Á., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence, 36*(6), 584-606. doi:10.1016/j.intell.2008.01.002

Conway, A. R. A., & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General, 123*(4), 354-373. doi:10.1037/0096-3445.123.4.354

Conway, A. R. A., Getz, S. J., Macnamara, B., & Engel de Abreu, P. M. J. (2011). Working memory and intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 394-418). Cambridge, United Kingdom: Cambridge University Press.

Conway, A. R. A., Kane, M. J., Bunting, M., Hambrick, D., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769-786. doi:10.3758/bf03196772

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7*(12), 547-552. doi:10.1016/j.tics.2003.10.005

Cowan, N. (1993). Activation, attention, and short-term memory. *Memory & Cognition, 21*(2), 162-167. doi:10.3758/bf03202728

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(01), 87-114. doi:10.1017/S0140525X01003922

Cowan, N. (2010a). The magical mystery four. *Current Directions in Psychological Science, 19*(1), 51-57. doi:10.1177/0963721409359277

Cowan, N. (2010b). Multiple Concurrent Thoughts: The Meaning and Developmental Neuropsychology of Working Memory. *Developmental Neuropsychology, 35*(5), 447-474. doi:10.1080/87565641.2010.494985

Cowan, N., Saults, J. S., Elliott, E. M., & Moreno, M. V. (2002). Deconfounding serial recall. *Journal of Memory and Language, 46*(1), 153-177. doi:10.1006/jmla.2001.2805

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466. doi:10.1016/s0022-5371(80)90312-6

Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review, 112*(1), 3-42. doi:10.1037/0033-295x.112.1.3

Doidge, N. (2010). *The brain that changes itself*. Melbourne: Scribe.

East of the Web. (2011, n.d.). definetime Retrieved February 20, 2012, from http://www.eastoftheweb.com/games/DefineTime1.html

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*(3), 309-331. doi:10.1037/0096-3445.128.3.309

Faul, F., Erdfelder, E., & Lang, A. (2010, January 1). G*Power3 Retrieved February 20, 2012, from http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, California: Sage.

Garavan, H. (1998). Serial attention within working memory. *Memory & Cognition, 26*(2), 263-276. doi:10.3758/bf03201138

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience, 6*(3), 316-322. doi:10.1038/nn1014

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: a new hypothesis. *Trends in Cognitive Sciences, 11*(6), 236-242. doi:10.1016/j.tics.2007.04.001

Hoskinson, P. (2008). Brain workshop - a dual n-back game Retrieved 20 February, 2012, from http://brainworkshop.sourceforge.net/

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America, 105*(19), 6829-6833.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences of the United States of America, 108*(25), 10081-10086.

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

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning — implications for training and transfer. *Intelligence, 38*(6), 625-635. doi:10.1016/j.intell.2010.09.001

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the n-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 615-622. doi:10.1037/0278-7393.33.3.615

Kane, M. J., & Engle, R. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review, 9*(4), 637-671. doi:10.3758/bf03196323

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*(2), 189-217. doi:10.1037/0096-3445.133.2.189

Kienle, G. S., & Kiene, H. (1997). The powerful placebo effect: fact or fiction? *Journal of clinical epidemiology, 50*(12), 1311-1318.

Li, S.-C., Schmiedek, F., Huxhold, O., Röcke, C., Smith, J., & Lindenberger, U. (2008). Working memory plasticity in old age: Practice gain, transfer, and maintenance. *Psychology and Aging, 23*(4), 731-742. doi:10.1037/a0014343

Martínez, K., Burgaleta, M., Román, F. J., Escorial, S., Shih, P. C., Quiroga, M. Á., & Colom, R. (2011). Can fluid intelligence be reduced to ‘simple’ short-term storage? *Intelligence, 39*(6), 473-480. doi: 10.1016/j.intell.2011.09.001

Mather, N., & Woodcock-Johnson, R. W. (2001). *Examiner’s manual: Woodcock-Johnson III tests of cognitive abilities*. Itasca, IL: Riverside Publishing Company.

McGrew, K. S., & Woodcock-Johnson, R. W. (2001). *Technical manual: Woodcock Johnson III*. Itsaca, IL: Riverside Publishing.

Mesulam, M. (2000). *Principles of behavioral and cognitive neurology*. New York: Oxford University Press.

Oberauer, K. (2005). Binding and Inhibition in Working Memory: Individual and Age Differences in Short-Term Recognition. *Journal of Experimental Psychology: General, 134*(3), 368-387. doi:10.1037/0096-3445.134.3.368

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25*(1), 46-59. doi:10.1002/hbm.20131

Palmer, V. (2011). *Improving intelligence through training on the single n-back task.* Unpublished hounour's thesis, Edith Cowan University, Joondalup.

Prandl, A. (2012). *Visual memory improvement in recognition*. Manuscript in preparation. Faculty of Computing, Health and Science. Edith Cowan Univerity. Joodalup.

Preece, D. (2011). *The effect of working memory (n-back) training on fluid intelligence.* Unpublished hounour's thesis, Edith Cowan University, Joondalup.

Raven, J., Raven, J. C., & Court, J. H. (2003). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 1: General overview*. San Antonio, TX: Harcourt Assessment.

Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: the first 345 studies. *Behavioral and Brain Sciences, 1*, 377-386. doi:10.1017/S0140525X00075506

Ross, S. M., & Di Vesta, F. J. (1976). Oral summary as a review strategy for enhancing recall of textual material. *Journal of Educational Psychology, 68*(6), 689-695.

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: Findings from the COGITO Study. *Frontiers in Aging Neuroscience, 2*, 1-10. doi:10.3389/fnagi.2010.00027

Sedgwick, P. (2012). The Hawthorne effect. *BMJ, 344*. doi:10.1136/bmj.d8262

Seidler, R. D., Bernard, J. A., Buschkuehl, M., Jaeggi, S. M., Jonides, J., & Humfleet, J. (2010). Cognitive training as an intervention to improve driving ability in older adults *Technical Report No. M-CASTL 2010-01*. Ann Arbor: University of Michigan.

Shelton, J. T., Elliott, E. M., Hill, B. D., Calamia, M. R., & Gouvier, W. D. (2009). A comparison of laboratory and clinical working memory tests and their prediction of fluid intelligence. *Intelligence, 37*(3), 283-293. doi:10.1016/j.intell.2008.11.005

Shelton, J. T., Metzger, R. L., & Elliott, E. M. (2007). A group-administered lag task as a measure of working memory. *Behavior Research Methods, 39*(3), 482-493. doi:10.3758/BF03193017

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin, 138*(4), 628-654. doi:10.1037/a0027473

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsycholigical test: Administration, norms and commentary* (2 ed.). New York, New York: Oxford Univeristy Press.

St Clair-Thompson, H. L. (2009). Backwards digit recall: A measure of short-term memory or working memory? *European Journal of Cognitive Psychology, 22*(2), 286-296. doi:10.1080/09541440902771299

Sternberg, R. J. (2008). Increasing fluid intelligence is possible after all. *Proceedings of the National Academy of Sciences of the United States of America, 105*(19), 6791-6792.

Tewes, U. (1991). *Hamburg-Wechsler-Intelligenztest für Erwachsene (HAWIE-R)*. Bern, Switzerland: Huber.

Tulving, E., & Osler, S. (1968). Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology, 77*(4), 593-601.

Tulving, E., & Patterson, R. D. (1968). Functional units and retrieval processes in free recall. *Journal of Experimental Psychology, 77*(2), 239-248.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127-154. doi:10.1016/0749-596x(89)90040-5

Unsworth, N., & Engle, R. W. (2007a). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*(1), 104-132. doi:10.1037/0033-295x.114.1.104

Unsworth, N., & Engle, R. W. (2007b). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin, 133*(6), 1038-1066. doi:10.1037/0033-2909.133.6.1038

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language, 62*(4), 392-406. doi:10.1016/j.jml.2010.02.001

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2010). The contributions of primary and secondary memory to working memory capacity: An individual differences analysis of immediate free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(1), 240-247. doi:10.1037/a0017739

Wechsler, D., Coalson, D. L., & Raiford, S. E. (2008). *WAIS-IV: Technical and interpretive manual*. San Antonio, TX: Pearson.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*(3), 441-517. doi:10.1006/jmla.2002.2864

Appendix A

Participant Information Letter, Participant Informed Consent, Participant Demographic Information, and Participant Feedback Forms

Participant Information Letter

Information Letter to Participants (On ECU letterhead)

Improving Intelligence through Cognitive Training

Thank you for considering participation in our research project. This study is a requirement for the Psychology Honours program at Edith Cowan University, which has been given approval by the Faculty of Computing, Health, and Science Ethics Sub-Committee.

The aim of this research is to investigate whether intelligence can be improved through cognitive training. The previous understanding of intelligence was that it was relatively fixed, with heredity being the major determinant. However, a number of recent studies have indicated that intelligence can be modified through specific cognitive training tasks.

As a participant you would be asked to:

1. Undertake three intelligence tests at a convenient day/time, at a convenient location of your choice. This process will take approximately 60 minutes.

2. Train using a computer based cognitive task, accessed through your own personal computer. The daily requirements of the training schedule are 18-20 minutes per day for 20 days once the initial intelligence testing has been completed.

3. Retake the initial intelligence tests. Again this should take about 60 minutes, and ideally would occur within three days of completing the cognitive training task.

Note: The schedule for the training task is flexible, however learning is maximised when there is some form of routine. The only stipulation is that the 20 days of training occur within a 30 day period. A recommended training schedule is 5 consecutive days, followed by two days break, which is equivalent to weekly training with the weekends off.

Participant benefits: Potential gains include an improvement in intelligence. Further, research has indicated that any gains attained appear to have longevity. Also access to the cognitive training task will remain, if continued practice is of interest.

The intelligence tests used in this research are similar to those which are used in recruitment and education. Therefore exposure to these tests may provide an advantage in future vocational testing or recruitment selection processes.

Feedback will be given with regard to the initial testing, providing a ranking of scores with similarly aged peers.

Potential risks / discomfort to participants: There are no foreseeable risks to participants in this study, apart from the inconvenience of committing to training and testing.

Confidentiality / Use of study data: All data gathered during the course of this study will remain confidential, and will not be disclosed to anyone outside of the research team. Part of the data collation process is the removal of all personal information, with names replaced with an alphanumeric identifier e.g. A23, and therefore no individual will be able to be personally identified with any of the data. There is a likelihood that the results will be published for scientific purposes, however using the above process ensures that individuals will maintain anonymity, and not be personally identifiable.

Choice to participate in the study: There is no mandatory obligation to participate in this study. No punishment, consequences or loss of benefits will occur should you choose not to participate.

Also if you elect to participate in the study, you are free to withdraw at any time without explanation, simply by contacting Paul Beavon, Allison Prandl, Dr Ken Robinson or Dr Ricks Allan at your earliest convenience.

Contacts: If you would like to take part in this project or require further information, please contact Paul Beavon or Allison Prandl via email or phone - contact details can be found below.

Thank you for your consideration,

Yours sincerely,

Paul Beavon Allison Prandl

Contact Details (for further information)	
Investigator Email Phone	Paul Beavon pbeavon@our.ecu.edu.au [redacted]
Investigator Email Phone	Allison Prandl aprandl@our.ecu.edu.au [redacted]
Supervisor Email Phone School Faculty	Dr Ken Robinson k.robinson@ecu.edu.au (08) 6304 5226 Edith Cowan University (Joondalup) School of Psychology and Social Science
Supervisor Email Phone School Faculty	Dr Ricks Allan m.allan@ecu.edu.au (08) 6304 5048 Edith Cowan University (Joondalup) School of Psychology and Social Science
Independent researcher Email Phone	Dr Andrew Guilfoyle a.guilfoyle@our.ecu.edu.au 6304 5192

Participant Informed Consent Form

Participant Informed Consent Form (On ECU letterhead)

Project Title: Improving Intelligence through Cognitive Training

Consent:

The signing of this form indicates that you have read the information letter provided and are interested in participating in the study.

I have been provided with the 'Information Letter to Participants' which I have read and understood. I am aware of the purpose of this study, my requirements as a participant and how the resulting data will be used. In accordance with the information provided:

- *I am volunteering to be a participant in 'Improving Intelligence through Cognitive Training' study*
- *I understand that I can withdraw from the study at any time without providing a reason*
- *I give permission for the data to be published without any of my personal details thereby maintaining my anonymity.*

Participants Name: ____________________________________

Participant's Signature: ____________________________________ Date: / / 2012

Researcher's Name: ____________________________________

Researcher's Signature: ____________________________________ Date: / / 2012

Participant Demographic Information Form

Participant Demographic Information Form

Name : ____

Age (years) : ___________

Gender : M F

Contact number : _________________________

Email address : ____

Study Identifier : _____________ (provided by researcher)

What is your highest qualification attained?

- o < Year 12
- o Year 12
- o University degree
- o Other qualification after Year 12

Are you currently studying at:

- o University
- o Another institution (TAFE, other colleges)

Participant Feedback Form

Improving Working Memory through Cognitive Training

Dear

Thank you for participating in the '*Improving Working Memory through Cognitive Training Study'.* Below is feedback regarding your performance on the tests you completed during the study.

Working Memory Capacity **Woodcock-Johnson III, Tests 7 & 9**	**Level of Performance**
Tests 7 & 9 of the WJIII evaluate short term memory span and auditory working memory. Short term memory refers to how many bits of information you can hold in your immediate memory while manipulating that information. Auditory working memory refers to short term memory stimulated by sound.	Your Working Memory score placed you at the __ percentile, which means that you performed as well as or better than __ of your age related peers.
Visual Working Memory **Woodcock-Johnson III, Test 13**	**Level of Performance**
Test 13 of the WJIII evaluates visual working memory, which refers to short term memory stimulated by vision.	Your Visual Memory score placed you at the __ percentile, which means that you performed as well as or better than __ of your age related peers.

If you have any questions about your scores please contact any of the following by phone or email:

Contact Details	
Investigator	Paul Beavon
Email	pbeavon@our.ecu.edu.au
Phone	[redacted]
Investigator	Allison Prandl
Email	aprandl@our.ecu.edu.au
Phone	[redacted]
Supervisor	Dr Ken Robinson
Email	k.robinson@ecu.edu.au
Phone	(08) 6304 5226
School	Edith Cowan University (Joondalup)
Faculty	School of Psychology and Social Science
Supervisor	Dr Ricks Allan
Email	m.allan@ecu.edu.au
Phone	(08) 6304 5048
School	Edith Cowan University (Joondalup)
Faculty	School of Psychology and Social Science
Independent researcher	Dr Andrew Guilfoyle
Email	a.guilfoyle@our.ecu.edu.au
Phone	6304 5192

Appendix B

Instructions for Configuring Brain Workshop Application for Single n-Back

1. Install as per website instructions http://brainworkshop.sourceforge.net/
2. Access "C:\Program Files\Brainworkshop\" through "My Computer"
3. <Double click> on brainworkshop.exe
4. <C> for configure
5. <Tab> down to "Audio"
6. <Space bar> to set "Audio" to off
7. <Enter> to "Apply" settings
8. Play one round of position *n*-back to save settings (failure to do this will cause settings to regress to dual *n*-back)
9. Access "C:\Program Files\Brain Workshop\data\" through "My Computer"
10. <Double click> config.ini to open with the text editor
11. Change the following settings:
    - JAEGGI_MODE = True
    - JAEGGI_SCORING = True
    - JAEGGI_FORCE_OPTIONS = True [default]
    - JAEGGI_FORCE_OPTIONS_ADDITIONAL = True [default]
    - JAEGGI_FALLBACK = 70
    - USE_SESSION_FEEDBACK = False
    - CHANCE_OF_GUARANTEED_MATCH = 0.2
    - DEFAULT_CHANCE_OF_INTERFERENCE = 0.1
12. Save settings through by clicking <File> <Save>

Appendix C

Contents of Installation CD

Folder	Description
Scoring sheets	Scoring sheet forms in PDF
Task 1	Single *n*-back task software
Task 2	Definetime and Millionaire URL links
Config	Brain Workshop single *n*-back customised configuration file
Copy_software	.bat file to copy software images to subject’s personal computer
Installation instructions	Instructions to copy software onto subject’s personal computer