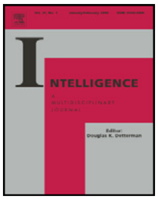
Intelligence 40 (2012) 531–542

Contents lists available at SciVerse ScienceDirect

Intelligence

journal homepage:



Working memory training does not improve intelligence in healthy young adults

Weng-Tink Chooi [a,b,*], Lee A. Thompson [a]

[a] *Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA*
[b] *Advanced Medical and Dental Institute, Universiti Sains Malaysia, 1-8 (Lot 8), Persiaran Seksyen 4/1, Bandar Putra Bertam, 13200 Kepala Batas, Pulau Pinang, Malaysia*

ARTICLE INFO

Article history:
Received 10 August 2011
Received in revised form 27 July 2012
Accepted 27 July 2012
Available online xxxx

Keywords:
Cognitive training
Dual n-back
Active control group
Fluid intelligence
Young adults

ABSTRACT

Jaeggi and her colleagues claimed that they were able to improve fluid intelligence by training working memory. Subjects who trained their working memory on a dual n-back task for a period of time showed significant improvements in working memory span tasks and fluid intelligence tests such as the Raven's Progressive Matrices and the Bochumer Matrices Test after training compared to those without training. The current study aimed to replicate and extend the original study in a well-controlled experiment that could explain the cause or causes of such transfer if indeed the case. There were a total of 93 participants who completed the study, and they were assigned to one of three groups—passive control group, active control group and experimental group. Half of the participants were assigned to the 8-day condition and the other half to the 20-day condition. All participants completed a battery of tests at pre- and post-tests that consisted of short timed tests, a complex working memory span and a matrix reasoning task. Although participants' performance on the training task improved, results from the current study did not suggest any significant improvement in the mental abilities tested, especially fluid intelligence and working memory capacity, after training for 8 days or 20 days. This does not support the notion that increasing one's working memory capacity by training and practice could transfer to improvement on fluid intelligence as asserted by Jaeggi and her colleagues.

1. Introduction

Part of the nature versus nurture debate is the issue of the malleability of intelligence (Wahlsten, 1997)—can the environment modify intellectual ability? Interests and efforts to raise intelligence as well as other cognitive abilities have been around for more than a century (Spitz, 1986). The idea that it may be possible to manipulate intelligence has been very appealing to researchers in education and the behavioral sciences, and the large body of research focused on aspects of the treatment of intellectual impairment provides an excellent example of these efforts.

Long term intervention programs to improve intelligence such as Head Start and the Abecedarian Project have not been successful. At the completion of the Abecedarian Project, results showed that there was substantial improvement in IQ scores in the experimental group compared to the control group but the superior performance quickly decreased when the project ended (Spitz, 1986). A large number of studies involving short-term interventions have also been conducted. In general, short-term intervention programs have not significantly improved latent ability but may have only increased task specific variance. These studies suggested that training or repeated practice on a task with instructional aid improved performance on that specific task but rarely did the improvement transfer to other general cognitive abilities (Belmont & Butterfield, 1977; Ferrara, Brown, & Campione, 1986).

---

* Corresponding author at: Advanced Medical and Dental Institute, Universiti Sains Malaysia, 1-8 (Lot 8), Persiaran Seksyen 4/1, Bandar Putra Bertam, 13200 Kepala Batas, Pulau Pinang, Malaysia. Tel.: +60 4 562 2304; fax: +60 4 562 2468.
E-mail addresses: wengtink@amdi.usm.edu.my (W.-T. Chooi), lat@case.edu (L.A. Thompson).

0160-2896/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.intell.2012.07.004

The past several years have seen a proliferation of research on cognitive training, especially working memory training (Morrison & Chein, 2011). Many studies have reported training and transfer effects as a result of working memory, executive function, and attention type training, most of which were done in young children or the older adult population. Klingberg, Forssberg, and Westerberg (2002) and Klingberg et al. (2005) observed improvements in matrix reasoning tasks besides reduced inattentive symptoms in children with Attention Deficit Hyperactivity Disorder (ADHD). Others found improvements in fluid reasoning after training on working memory in the older adult population (Borella, Carretti, Riboldi, & De Beni, 2010; Schmiedek, Lovden, & Lindenberger, 2010), task switching training in three different age groups (Karbach & Kray, 2009), executive control/planning training in the older adult population (Basak, Boot, Voss, & Kramer, 2008) and attention training in children (Rueda, Rothbabrt, McCandliss, Saccomanno, & Posner, 2005). Studies done on old adults that employed working memory training (Buschkuehl et al., 2008) and strategic training (Carretti, Borella, & De Beni, 2007) impacted memory performance, a near transfer effect, and these studies did not report significant improvements on g_f tasks. Van der Molen, Van Luit, Van der Molen, Klugkist, and Jongmans (2010) found short-term memory improvement but no IQ improvements in adolescents with mild intellectual disability. Minear and Shah (2008) employed a task-switching training paradigm and did not report improvements on IQ. Li et al. (2008) reported no far transfer effects from working memory training to complex span tasks and did not report improvements in IQ.

There are plenty of studies that found near transfer after working memory, executive functions and attention training but not so much on far transfer effects. Near transfer refers to changes in a domain caused by changes in another similar domain due to comparable ability or process, and far transfer effects refer to changes in domains caused by changes in a separate domain of different processes. Many of these studies were conducted on children ranging from 4 to 11 years of age, when their cognitive abilities were still developing and have not reached maturity (Fry & Hale, 2000). Bergman-Nutley et al. (2011) reported consistent near transfer effects in their study on 4-year old children. The authors observed improvements on reasoning tasks in groups that trained on reasoning skills, and they did not find transfer effects from working memory training to reasoning or fluid intelligence tasks. More examples of near transfer effect include a study by Mackey, Hill, Stone, and Bunge (2011), St. Clair-Thompson, Stevens, Hunt, and Bolder (2010) and Thorell, Lindqvist, Bergman-Nutley, Bohlin, and Klingberg (2009) that reported near transfer effects but neither far transfer effects nor IQ improvements. Holmes, Gathercole, and Dunning (2009) reported that IQ scores were unaffected by working memory training in children with ADHD, and Holmes et al. (2010) reported no boost in IQ performance in children with low working memory capacity. A recent review by Diamond and Lee (2011) on cognitive training conducted on children concluded that only core executive function—working memory, cognitive flexibility and inhibition—training is most beneficial to 4–12 year-olds, and most studies cited in this review reported near transfer effects (Diamond & Lee, 2011).

Nearly all the studies mentioned above that found transfer effects were studies that were conducted on children and older adults, when their cognitive development or decline was relatively malleable than young adults (Borella et al., 2010; Fry & Hale, 2000). Modifying cognitive abilities did not seem to be difficult during periods of growth when intervention could facilitate and perhaps accelerate development and maturity (Rueda et al., 2005). Modification also seemed possible in old adults when their cognitive decline could be delayed with intervention (Basak et al., 2008; Borella et al., 2010; Buschkuehl et al., 2008), consistent with the "disuse" hypothesis (Orrell & Sahakian, 1995). This hypothesis has been supported by animal and human studies that demonstrated considerable neuronal plasticity due to increased activities from experiential input and perceptual-sensory stimuli (Tranter & Koutstaal, 2008). However, studies that suggested possible modification of cognitive abilities in young adults, when general cognitive abilities are less malleable compared to childhood and aging periods were considerably fewer (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Jaeggi et al., 2010; Karbach & Kray, 2009; Schmiedek et al., 2010). More studies should be carried out on healthy, young adults before claiming with confidence that general cognitive abilities such as fluid reasoning could be improved with short periods of cognitive training.

Jaeggi et al. (2008) reported that they had improved fluid intelligence (g_f) of young adults in a study by training their working memory (WM) through repeated practice with a dual N-back task. They argued that since WM and g_f shared common variance (Ackerman, Beier, & Boyle, 2005; Colom, Flores-Mendoza, & Rebollo, 2003; Fry & Hale, 1996; Jurden, 1995; Kane, Hambrick, & Conway, 2005; Kyllonen & Christal, 1990; Oberauer, Schulze, Wilhelm, & Süβ, 2005; Stauffer, Ree, & Carreta, 1996; Tucker & Warr, 1996; Verguts & De Boeck, 2002), engaging neural circuits shared by WM and g_f by training WM may transfer to improvements in g_f. Studies in cognitive psychology and neuroscience that tried to explain the relationship between WM and g_f such as Halford, Cowan, and Andrews (2007) suggested that WM and reasoning skills shared related capacity limits, and that the common thread between the two functions was the shared requirement to bind elements to slots of a hypothetical coordinate system in one's memory. The process of maintaining the bindings between elements required attention, which was essential to WM and reasoning abilities (Halford et al., 2007). Gray, Chabris, and Braver (2003) suggested that the relation between g_f and WM was mediated by activities in the lateral prefrontal and parietal regions. Kane and Engle (2002) reported that the dorsolateral prefrontal cortex could have a role in WM especially related to attention control. Conway, Kane, and Engle (2003) supported the hypothesis by Gray et al. (2003) that WM span tasks activate regions in the prefrontal cortex when the executive-control mechanism is recruited to combat interference during the maintenance and manipulation of information.

The assumption that WM and g_f may share the same neural network and mental resources was the theory behind Jaeggi et al.'s (2008) hypothesis that training to improve one's WM could transfer the improvement to g_f. The adaptive nature of the dual N-back task used for training was intended to engage the executive attention at all times so that automatic responses could not develop. It was suggested that under consistent format and information conditions, practice would lead to automatic responses and less mental or attentional resources would be employed. However, in variable information or inconsistent contexts, controlled processing that utilized mental resources would still have been taken place even after practicing on the task for a considerable period of time (Ackerman, 1987). Jaeggi et al. (2008) were essentially targeting participants' executive attention in their WM training with the implementation of the adaptive feature of the training task. Therefore, an increase in WM performance at the end of training could mean an increase in attention span. If the effects of WM could be quantified through the analysis of goal management (Carpenter, Just, & Shell, 1990), improvements in WM could mean better and improved ability to manage representations of information. These effects could be measured by improved performance on fluid ability tests, such as the Raven's, and mental rotation tests where steps of abstractly manipulating the movements of three-dimensional objects must be actively managed in one's mind.

After decades of unsuccessful and inconclusive research and efforts to raise intelligence, a study that suggested otherwise in a sample population of individuals when their general cognitive ability is strongly suggested to have matured (Cattell, 1987) should be subjected to further examination through replication. Moody (2009) identified some weaknesses in the study by Jaeggi et al. (2008), and one of them was participants in the experimental group did not take the same IQ test. Those who trained the least (8 training sessions) did the Raven's Advanced Progressive Matrices (RAPM) and the rest of the participants who trained either 12, 17 or 19 days took the Bochumer Matrices Test (BOMAT). The researchers in the study reported significant group differences for the 12-day, 17-day and 19-day training groups. Participants who performed the RAPM in the 8-day group did not show significant improvement in their IQ scores after training for 8 days (Jaeggi et al., 2008). The reason behind this observation could either be that longer training produced more score gains, or the nature of the training task itself facilitated better test taking specifically for the BOMAT (Moody, 2009).

Both the IQ tests—the RAPM and BOMAT—employed in the study by Jaeggi et al. (2008) shared some similarities. Each item in both tests was a matrix of figures and a spot in the matrix was left blank. Items on RAPM consisted of 3×3 matrices, while items on BOMAT consisted of 5×3 matrices. The figures in the matrices were arranged according to a pattern, and test-takers would have to identify the pattern, or patterns, unique to each matrix in order to infer a solution from multiple possible answers given at the bottom of the matrices. Another similarity between the two tests was that the difficulty level of the questions increased as the test-taker progressed through either test. A high score would reflect the test-taker's ability to solve the difficult items. Test-takers would also be able to learn how to solve subsequent questions that were progressively harder based on the patterns they inferred from previous items that they had solved. Participants in the Jaeggi et al. (2008) study were not given the opportunity to attempt the more difficult questions because the researchers essentially removed the progressive nature of the tests by reducing the allotted time to take the test from 45 min to 10 min (Moody, 2009). When Jaeggi et al. (2010) replicated their results, they defended their time constraint testing protocol by citing studies by Salthouse (1993) and Unsworth and Engle (2005) that suggested no evidence for differential working memory effects for the various items on the RAPM. Jaeggi and her colleagues argued that these studies (Salthouse, 1993; Unsworth & Engle, 2005) provided justification that limiting participants to the first several items on the RAPM and BOMAT does not affect their measure of G_f (Jaeggi et al., 2010).

The present study aimed to replicate and extend the findings reported by Jaeggi et al. (2008), in addition to correcting the potential confounds identified by Moody (2009) and comparing results of WM training with well-controlled groups. In other words, the treatment group should differ from the control group only by the very element that affected the outcome. The objective of the current study was to determine if effects from WM training specifically, a core executive function (Diamond & Lee, 2011), would transfer to improvement in g and/or g_f (producing a far transfer effect) in young adults. Though Jaeggi et al. (2008, 2010) specifically targeted g_f, this construct was suggested to be inadequate to capture the general intellectual ability, g (Johnson & Bouchard, 2005a). A hierarchical model of human intelligence has been widely accepted in the field of intelligence (Deary, 2001). The three-strata model of human intellectual abilities proposed by Carroll in 1993 (Carroll, 2003), and the practice of structural equation modeling in the field have been providing converging support that human intelligence could be described in terms of a hierarchical structure. Johnson and Bouchard (2005a) suggested a model of general intelligence, g, which influenced three factors—verbal, perceptual and mental rotation (VPR). The VPR model proposed has been repeatedly and empirically tested (Johnson & Bouchard, 2005a, 2005b; Johnson, te Nijenhuis, & Bouchard, 2007) and offered an excellent example of the current direction and general acceptance of a hierarchical organization of human intelligence; thus this model would be applied in the present study as a measure of g using a theoretically based, comprehensive battery of cognitive ability tests to provide insight into specific mental processes influenced and not influenced by WM training. As Jaeggi et al. (2008) stated, "… tasks that measure G_f are picking up other cognitive skills as well, and perhaps the training is having an effect on these skills even if measures of capacity are not sensitive to them" (p. 6831). If WM training could improve intellectual performance, the improvement may be due to global effects or specific effects. The present study included a series of tests that supported the model postulated by Johnson and Bouchard (2005a), and none of the tests used possessed features similar to those in the working memory task used for training. The present study predicted that there would be no improvements in verbal and perceptual tests, but there could be improvements in spatial ability and matrix reasoning tests.

2. Methods

Participants were students enrolled in Introduction to Psychology, Health Psychology or Quantitative Methods in Psychology at a private Midwestern university (N = 130). They were initially randomly assigned into one of 6 groups—2 experimental groups, 2 passive control groups and 2 active control groups. Participants assigned to the experimental or active control groups were given the option to skip the lab component completely and only return to post-test session only, thus they were allowed to reassign themselves to the passive control groups. Participants from Introduction to Psychology received 6 research credits for the first 6 h of participation and $7.50 for each subsequent hour in the study. Participants from Health Psychology and Quantitative Methods in Psychology received 4 extra credits for their participation in the pre- and post-test sessions and $7.50 for each subsequent hour they participated in the study. There were two conditions within each group—8-day and 20-day intervals between pre- and post-test sessions. These conditions differed slightly from the original study (Jaeggi et al., 2008), in which the four training durations were 8 days, 12 days, 17 days and 19 days. In the original study, there were no significant improvements in the 8-day condition but significant changes were detected in the 17-day and 19-day conditions; therefore the researchers claimed that improvement was dosage-related—more training led to more gains (Jaeggi et al., 2008). The current study compared only two conditions, an 8-day and 20-day training period. Every participant completed a pre- and post-test session before and after training to assess ‘g’ and specific cognitive abilities with the same test battery. Participants in the active control group were not told that they were actually part of a control group, and they were under the impression that they were in the experimental group undergoing WM training. Participants trained once a day (for about 30 min), four days a week. Training schedule was determined by participants, and they scheduled training appointments when they decided to participate in the training part, before training sessions began. Participants completed their training within the restricted time frame of 8 or 20 days. Post-test was administered the day after the last training session, and the gap between training and post-test was similar and comparable across groups.

The present study incorporated two separate control groups to control for simple practice effects and/or the Hawthorne effect (increased motivation simply because of the attention paid to the participants' task performance). The first control group in the present study controlled for simple practice effects as did Jaeggi et al. (2008). The second (active) control group was exposed to the same experimental format as the first control group with the addition of a filler task designed to be comparable to the training task without its complexity. Participants assigned to the second control group trained for the same amount of time as their counterparts in the experimental training group. The filler task had the same modality as the training task but without the adaptive feature inherent in the training task that purportedly contributed to the training effects. The adaptive nature of the training task was alleged to contribute to the improvement of working memory capacity and performance in intellectual abilities (Jaeggi et al., 2008), so this allegation could be verified if there were significant differences between the experimental group and both of the control groups.

The training task employed in the present study was the dual n-back task (Jaeggi et al., 2008). In this computerized task, a series of blue squares was flashed on the screen at one of eight different locations in random order. Simultaneously, a series of letters was presented through an audio output. Each visual and auditory stimulus was presented for 500 ms, and there was an interval of 2500 ms between stimuli. Participants were asked to press the “A” key when the location of the current square presented matched the location N stimuli back and the “L” key when the letter currently presented was the same as the letter presented N stimuli back. If both visual and audio stimuli presented were identical to the ones presented N stimuli before, participants press both “A” and “L” keys (Fig. 1). The N value was the same for both streams of stimuli. The task was designed to adapt to the user's performance so that the task would remain challenging to the user. If the users performed adequately in the current N-back task, i.e., they correctly identified matched stimuli in both modalities more than 90% of the time, the program would increase the difficulty level to $N+1$. However, if the user's performance dropped below 70% the program would reduce the difficulty level to $N-1$. The difficulty level remained unchanged if users did not meet either condition. The ability of the task to adapt to users' performance so that it remained demanding to the user was purported to have increased working memory capacity that translated to improved measured fluid intelligence (Jaeggi et al., 2008).

The active control task was a modified web-based version of the training task—the dual n-back task—publicly available for web users. In the modified version, instead of the program adapting automatically to the user's performance, the user only worked on a fixed level of difficulty. In the study, participants in the active control group working on the filler task completed 20 trials of the dual 1-back working memory task each time they came in for their “training” session.

The tests administered in the pre- and post-tests were the Mill-Hill vocabulary test, vocabulary tests (parts I and II) from the Primary Mental Abilities test battery, Word Beginning and Ending test, Colorado Perceptual Speed Test, Identical Pictures, Finding A's, Card Rotation and Paper Folding from the ETS test battery, Shepard–Metzler Rotation Test (1971) and Raven's Advanced Progressive Matrices (1990). These tests were chosen based on the tests that had among the highest primary loadings on the VPR model presented by Johnson and Bouchard (2005a). For every test described below except for Card Rotation, test scores were determined by the number of items answered correctly. The maximum score for these tests would be the number of items of each test.

The *Mill-Hill* vocabulary test was a multiple choice test where subjects chose the option word that was synonymous to the target word. There were 22 items in each part.

The *Vocabulary* test consisted of target words that had similar meanings to one of the four words given as options. Test-takers must identify the correct answer by circling the word that was synonymous or had the closest meaning to the target word. There were 50 items in part I and 25 items in part II. See Results section for an explanation of the inconsistency.

In the *Word Beginning and Ending* test, subjects were asked to generate as many words as they could that began and ended with letters specified by the test.

The *Colorado Perceptual Speed Test* (multiple-choice) was a test where subjects chose the exact copy of the target group of characters, for instance, ‘vgef’ or ‘9c6d’. There were 30 items in each part.

Identical Pictures shared a similar format but instead of characters as target stimuli, this test utilized pictorial objects as target. There were 48 items in each part.

The *Finding A's* test consisted of multiple columns of words, and subjects were asked to identify and cross out words that contain the letter ‘a’. There were 820 words with 100 words that contain the letter ‘a’ in each part.

In *Card Rotations*, a target shape was presented next to eight versions of the target. These versions were either rotated or flipped or both. Subjects must identify if each of the eight rotated shapes was on the same side or the flipped side of the target shape. There were 14 items in each part. This test was scored by the number of correctly identified rotation minus the number of incorrectly identified rotation. The maximum score

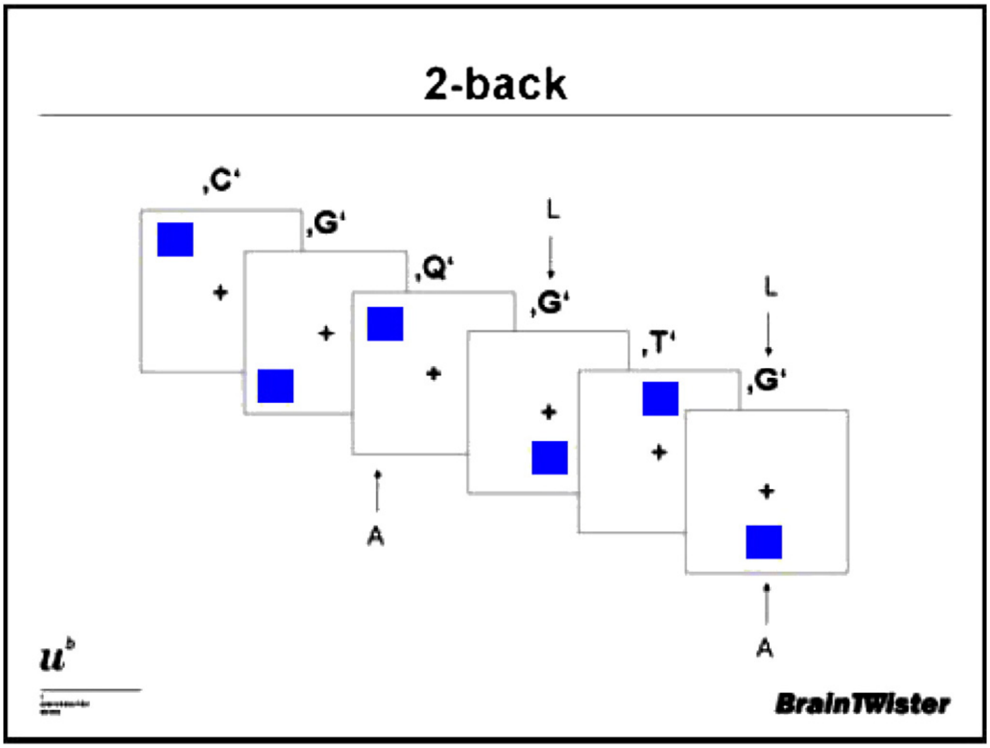


Fig. 1. An example of the instruction page (modified from the task's screenshot) that participants viewed prior to each block. In this figure, the block that was to be administered when participants proceeded from this page would be a 2-back task. Participants would need to respond if the current stimuli presented match the stimuli 2 trials back.

for this test would be 112 (14 items multiplied by 8 correct comparisons).

In *Paper Folding* (multiple choice), an illustrated piece of paper was shown to be folded in certain ways. After the paper was folded, a hole was punched at a certain location. Subjects must then visualize where the holes would be after the piece of paper was unfolded back to its original condition. There were 10 items in each part.

The *Mental Rotation* was a visualization test of mental spatial manipulation. Participants were asked to compare images of rotated 3-dimensional blocks to a target figure, and identify the two images of the target figure that were rotated at different angles. There were 10 items in each part.

There were 36 items in the *Raven's Progressive Advanced Matrices* (*RAPM*) and the difficulty level increased for subsequent questions; hence the descriptor "progressive" in its name. Each question consisted of a 3 × 3 matrix with the last entry of the matrix left blank. The matrices were formed based on a certain pattern, and the blank spot could be filled in with one of the eight options given once the pattern was identified.

All tests except for the Mill-Hill and RAPM were timed. Participants were given 90 s to complete as many items as they could in the Vocabulary tests, Word Beginning and Ending Test, Card Rotation, Paper Folding and Shepard–Metzler Mental Rotation tests. They were given 30 s to complete as many problems as they could on the Colorado Perceptual Speed Test and 45 s on the Identical Pictures test. The time limit was halved from the recommended time to accommodate the restriction of range in abilities from the participants in our sample, who were all college students, to avoid ceiling effects. Parts I and II of these tests were administered in the pre- and post-tests respectively. The odd number questions of the RAPM were administered in the pre-test and the even number questions in the post-test.

A WM span task was included in the pre- and post-test sessions. The Operation Span (OSPAN) was administered in group format using Microsoft Power Point that was designed by Shelton, Metzger, and Elliott (2007). This task required participants to evaluate a simple mathematical equation and remember a word that appears after the equation. There were four to seven equations and to-be-remembered (TBR) words for each trial. At the end of each trial, a slide would appear to prompt participants to list all the words shown in the previous trial. Each slide that contained the math equation and TBR word appeared for 8 s and participants were given 15 s to list all the words. There was a chimed sound reminder to warn participants that the next trial would begin in 5 s.

At pre-test, participants first completed all questionnaires before they began the group-administered OSPAN (reliability, $r = 0.75$). Then the timed-tests were administered in the following order (test–retest reliability in parenthesis):

1. Vocabulary ($r = 0.67$)
2. Colorado Perceptual Speed Test ($r = 0.69$)
3. Paper Folding ($r = 0.72$)
4. Word Beginning and Ending ($r = 0.66$)
5. Identical Picture ($r = 0.80$)
6. Card Rotations ($r = 0.67$)
7. Finding A's ($r = 0.86$)
8. Mental Rotation ($r = 0.78$)

Table 1
Demographic information.

	Passive control		Active control		Training	
	8-day	20-day	8-day	20-day	8-day	20-day
N (females)	22 (16)	23 (9)	15 (6)	11 (11)	9 (8)	13 (10)
$N_{\text{originally assigned}}$	15	14	17	16	15	16
Mean SAT Scores	1335	1351	1334	1368	1327	1397
(SD)	(100.4) n = 17	(124.0) n = 21	(124.3) n = 15	(175.5) n = 11	(41.1) n = 4	(115.8) n = 11

Demographic information of participants broken down according to the group they were assigned to. $N_{\text{originally assigned}}$ referred to the number of participants that was originally assigned to each group. N reported in the first row indicated the actual number of participants in each group. The change in N was due to the option that participants had to switch groups.

After the completion of the timed tests, participants were allowed to complete the RAPM (r = 0.68) and Mill-Hill Vocabulary Scale (r = 0.79) with no time constraint. Post-test sessions were conducted in the same order of test administration at pre-test, except that there were no questionnaires to complete at post-test and the untimed Mill-Hill and RAPM were conducted before the series of timed tests.

3. Results

There were a total of 130 participants who came to pre-test sessions but only 93 of them came back for post-test sessions—60 of whom were females. The following results were obtained from analyzing data provided from the 93 students who completed both pre- and post-tests. They averaged slightly younger than 20 years of age. There were more participants in the passive control group compared to the active control or experimental group because those assigned to the latter two groups were given the option to withdraw from the portion where they were asked to participate in the lab portion for 8 or 20 days. Most who withdrew were willing to return to complete the post-test, and these participants were assigned to the passive control group. Independent samples t-test analysis showed no difference in age and SAT scores for those who returned to post-test and those who did not ($t(110) = -1.63$, $p = 0.11$). One-way Analysis of Variance (ANOVA) did not show any significant differences in SAT mean scores among the six groups, and independent sample t-test comparing scores from the 8-day against the 20-day condition also supported the results from one-way ANOVA ($F(5, 106) = 0.48$, $p = 0.79$), which suggested that the average participant in each group behaved similarly.

Table 2 presented raw scores of all test measures for each group at each session (pre- and post-tests) and Table 3 presented composite measures for each construct—verbal, perceptual and mental rotation. Note that raw scores for Vocabulary in Table 2 were inconsistent—there were 50 items in Part 1 and only 25 items in Part 2. There was a mistake in the administration of Vocabulary, where Part 1 should have been split into two parts for pre- and post-test administrations; thus, scores from Vocabulary were removed from subsequent analyses following another reviewer's suggestion. All the other tests were analyzed using their raw scores in subsequent analyses.

3.1. Gender differences

Some gender differences were observed. Male participants scored higher on the SAT-Math ($t(70) = 3.18$, $p < 0.01$), and they performed significantly better than their female counterparts on the Mental Rotation ($t(128) = 3.33$, $p < 0.01$ on pre-test and $t(90) = 3.04$, $p < 0.01$ on post-test) and Card Rotation pre-test ($t(124) = 2.92$, $p < 0.01$). Female participants performed significantly better on Word Beginning and Ending pre-test ($t(124) = -1.98$, $p = 0.05$) and Finding A's ($t(128) = -3.81$, $p < 0.01$ on pre-test and $t(92) = -2.87$, $p < 0.01$ on post-test). These differences will be taken into consideration in Analysis of Variance, where the interaction effects between gender and group will be explored.

3.2. Working memory training

There were a total of 22 participants in the working memory training groups, 13 of whom trained for 20 days. The average of their performance on the training task and variance among participants for each session during the 8 and 20 days of training are presented in Figs. 2 and 3 respectively. The figures suggest that most of the participants improved their performance on the task as they practiced on it, and the variance among all the participants tended to increase as the number of training periods increased. Some participants were able to make more improvements than others. The final training session and the first training session are highly correlated (r = 0.79, $p < 0.01$). Most students did not achieve their highest performance on the final training session so a variable that indicated each participant's highest training score was computed. This variable significantly correlated with performance on the final training session (r = 0.98, $p < 0.01$) and their first training score (r = 0.75, $p < 0.01$). Another variable, which measured how much participants had improved over the course of their training, was created by subtracting the first initial training score from the highest training score. This variable, named WM-Improve, will be used in Regression analysis to predict test variables for participants in the experimental group.

The percentage of improvement for each training condition was calculated by using the following formula:

$$\%\,Improvement = \frac{Avg.Highest\,Training\,score - Avg.First\,Training\,score}{Avg.Highest\,Training\,score} \times 100.$$

Participants in the 8-day condition had a 34% improvement and those in the 20-day condition improved by 44%.

Participants' first WM training score was significantly related to Card Rotation post-test ($p < 0.05$) Their highest WM training score (WM-High) and WM-Improve correlated significantly with Card Rotation post-test at 0.57 ($p < 0.01$) and 0.50 ($p < 0.05$) respectively. Similarly, these variables had significant correlations with RAPM post-test (WM-High: r = 0.50, $p < 0.01$; WM-Improve: r = 0.51, $p < 0.05$). WM-High correlated significantly with score gain in Card Rotation at 0.47 ($p < 0.05$) and gain in Finding A's at 0.44 ($p < 0.05$). Mean working memory performance was also significantly correlated with score gain in Card Rotation (8-day: r = 0.58, $p < 0.01$; 20-day: r = 0.52, $p < 0.05$).

Table 2
Raw scores for all tests.

Time	Measure	8 day training (N = 9)	20 day training (N = 13)	8 day active control (N = 15)	20 day active control (N = 11)	8 day passive control (N = 22)	20 day passive control (N = 23)
PRE	Mill-Hill	18.6 (*5.46*)	19.1 (*3.09*)	19.9 (*4.16*)	19.2 (*4.53*)	17.8 (*3.16*)	18.3 (*4.93*)
	Vocab	23.1 (*7.20*)	24.9 (*5.23*)	24.9 (*2.52*)	26.9 (*5.56*)	24.4 (*7.07*)	23.0 (*5.67*)
	WBE	8.2 (*2.91*)	9.8 (*3.10*)	8.0 (*4.02*)	10.0 (*4.29*)	9.1 (*4.36*)	7.8 (*3.30*)
POST	Mill-Hill	19.4 (*4.45*)	18.4 (*1.85*)	17.5 (*2.98*)	19.1 (*3.56*)	17.5 (*3.12*)	18.4 (*3.87*)
	Vocab	8.9 (*6.21*)	8.4 (*3.20*)	8.5 (*3.56*)	7.7 (*3.87*)	7.5 (*3.70*)	7.7 (*3.34*)
	WBE	7.8 (*3.19*)	7.9 (*3.23*)	7.5 (*2.85*)	8.2 (*3.25*)	7.1 (*2.64*)	6.1 (*2.03*)
PRE	CPST	13.0 (*2.33*)	14.6 (*1.97*)	12.9 (*1.86*)	10.3 (*2.63*)	13.3 (*2.00*)	13.5 (*1.90*)
	Identical Picture	20.3 (*4.27*)	21.3 (*3.59*)	19.7 (*4.58*)	20.1 (*2.85*)	20.7 (*5.83*)	20.0 (*4.02*)
	Finding A's	15.3 (*3.00*)	15.4 (*2.96*)	13.8 (*3.34*)	14.9 (*4.06*)	16.8 (*3.10*)	15.5 (*4.21*)
POST	CPST	11.3 (*2.80*)	11.4 (*2.31*)	10.4 (*1.60*)	10.3 (*2.63*)	10.4 (*2.27*)	10.8 (*1.84*)
	Identical Picture	17.7 (*3.51*)	18.8 (*2.87*)	17.0 (*3.80*)	15.5 (*3.78*)	17.2 (*5.56*)	17.6 (*3.86*)
	Finding A's	14.3 (*2.34*)	14.8 (*3.56*)	14.0 (*2.85*)	14.6 (*3.41*)	14.5 (*2.67*)	13.7 (*2.99*)
PRE	Paper Folding	5.3 (*1.55*)	4.0 (*1.23*)	5.7 (*1.98*)	4.4 (*2.37*)	4.9 (*2.63*)	5.6 (*2.56*)
	Card Rotation	28.4 (*8.70*)	35.8 (*9.62*)	30.6 (*15.00*)	29.4 (*11.78*)	28.1 (*15.63*)	30.6 (*13.72*)
	Mental Rotation	3.4 (*1.94*)	2.0 (*1.35*)	3.7 (*1.28*)	2.4 (*1.97*)	1.9 (*1.63*)	3.0 (*1.85*)
POST	Paper Folding	6.2 (*0.95*)	4.5 (*1.66*)	5.3 (*1.24*)	5.4 (*2.07*)	4.8 (*1.70*)	5.7 (*1.68*)
	Card Rotation	35.8 (*10.97*)	34.8 (*12.12*)	34.9 (*9.97*)	31.9 (*11.71*)	29.9 (*12.70*)	36.4 (*11.03*)
	Mental Rotation	1.7 (*1.12*)	2.1 (*1.38*)	2.5 (*1.96*)	2.3 (*1.25*)	1.9 (*1.86*)	2.5 (*1.50*)
PRE	OSPAN	55.7 (*7.09*)	57.5 (*6.36*)	58.8 (*4.82*)	58.4 (*7.26*)	56.5 (*6.79*)	57.3 (*5.38*)
	RAPM	12.8 (*1.79*)	12.2 (*2.05*)	13.6 (*1.91*)	14.2 (*3.09*)	12.4 (*2.64*)	12.4 (*3.02*)
POST	OSPAN	56.7 (*5.89*)	58.1 (*5.87*)	58.7 (*4.46*)	58.2 (*8.75*)	56.5 (*6.54*)	57.4 (*6.34*)
	RAPM	12.7 (*2.00*)	12.1 (*2.81*)	13.3 (*1.91*)	13.4 (*2.70*)	11.3 (*2.59*)	11.9 (*2.64*)

Raw scores for all tests administered for each group in pre- and post-test sessions. WBE = Word Beginning and Ending; CPST = Colorado Perceptual Speed Test; OSPAN = Operation Span; RAPM = Raven's Advanced Progressive Matrices.

3.3. Paired t-tests

Paired t-test analyses revealed several trends. First, working memory capacity and intellectual ability as measured by Raven's Advanced Progressive Matrices (RAPM) did not show significant improvement in post-test performance compared to pre-test. Second, there was also no significant improvement in the spatial ability tests. Third, data suggested worse performances on verbal fluency and perceptual speed tests; however, a comparison of data in a pilot study where data of all test variables in both pre- and post-tests was collected from the same administration instead of two separate administrations at different times showed similar trend. This trend seemed to indicate that Part 2 of these tests, which were used in the post-test, were more difficult than Part 1 of these tests. Effect sizes (Cohen's *d*) for each test variable are presented in Table 4.

3.4. ANOVA

Analysis of Variance (ANOVA) was conducted using gain scores (post-test minus pre-test scores), and these analyses provided additional support to results obtained from paired t-tests. Particularly, in the 8-day condition only the perceptual speed test Finding A's, $F(2, 43) = 4.09$, $p = 0.02$, had significant changes across the three groups. Further investigation using ANOVA planned comparisons revealed that participants in the active control group had the highest gain which is significantly different from the other two groups. In the 20-day condition, one-way ANOVA did not show any significant changes in gain scores for the test variables.

3 × 2 factorial ANOVA was conducted on the gain scores for each test variable, with the group (controls vs. experimental) and condition (8-day vs. 20-day) as independent variables.

Table 3
Composite measures for verbal, perceptual and rotation constructs.

Time	Construct	8 day training (N=9)	20 day training (N=13)	8 day active control (N=15)	20 day active control (N=11)	8 day passive control (N=22)	20 day passive control (N=23)
Verbal	Pre	38.35	41.35	40.35	42.65	39.1	37.6
	Post	36.1	34.7	33.5	35.0	32.1	32.2
Perceptual	Pre	48.6	51.3	46.4	45.3	50.8	49.0
	Post	43.3	45.0	41.4	40.4	42.1	42.1
(Mental) rotation	Pre	37.1	41.8	40.0	36.2	34.9	39.2
	Post	43.7	41.4	42.7	39.6	36.6	44.6

Scores on Vocabulary in pre-test were halved to match Vocabulary scores in post-test before added up with scores from other verbal fluency tests. There were 50 items administered in pre-test and 25 items in post-test.

There were neither significant main effects nor interaction effects produced by the analysis.

Two-way ANOVA was conducted to test for gender effects and their interaction with group effects for tests that displayed gender effects as reported earlier, such as Mental Rotation, Card Rotation and Finding A's. This analysis showed that there were no significant interaction effects between gender and group.

3.5. ANCOVA

Analysis of Covariance (ANCOVA) was conducted to determine if the outcome at post-test was influenced by one's initial ability as measured at pre-test. One-way ANOVA was conducted on all the pre-test variables before conducting an Analysis of Covariance (ANCOVA). This analysis showed that there were differences in initial level ability across groups for the Mental Rotation test, $F(5, 87) = 3.16$, $p = 0.011$. The existing differences at pre-test did not allow for ANCOVA for this variable.

In the 8-day condition, all the pre-test variables significantly predicted the outcome on all corresponding post-test variables. Paper Folding showed a significant difference across groups after controlling for pre-test performance, $F(2, 41) = 3.51$, $p < 0.05$, *partial* $\eta^2 = 0.15$, and so did Raven's, $F(2, 39) = 3.89$, $p < 0.05$, *partial* $\eta^2 = 0.16$, after controlling for pre-test differences. In the 20-day condition, all the pre-test variables significantly predicted the outcome on all corresponding post-test variables except for the Card Rotation test.

3.6. Regression

A simple regression with improvement in N-back task as predictor was conducted using data from participants in the WM training group. This predictor did not significantly contribute to any variance in gain scores for all the variables in pre- and post-tests.

4. Discussion

Results from the current study did not suggest improvement in general intelligence after repeated training on a challenging working memory task. Our prediction that spatial and reasoning abilities could be improved after working memory training was not supported. Paired t-tests, Analysis of Variance (ANOVA) and Analysis of Covariance (ANCOVA) did not show any significant changes between pre- and post-test scores across the experimental and control groups. This observation is in contrast to the conclusion published by Jaeggi et al. (2008) and replicated recently in 2010. Participants who trained their working memory in the current study appeared to improve on the training task just as much those in the original study (Jaeggi et al., 2008), yet they showed no significant improvement in solving items on the RAPM even after training for 20 days. From the data that they published, it can be estimated that participants in the 8-day training condition improved by 34% and those in the 19-day condition improved by 47% (Jaeggi et al., 2008). Participants from the current study

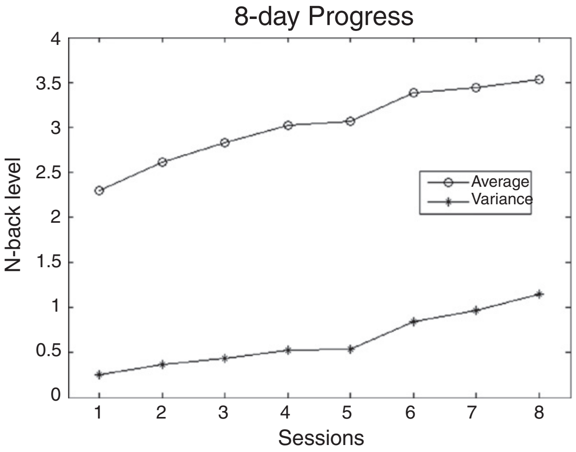


Fig. 2. Average performance on the training task and variance among all the participants in the first 8 days of training.

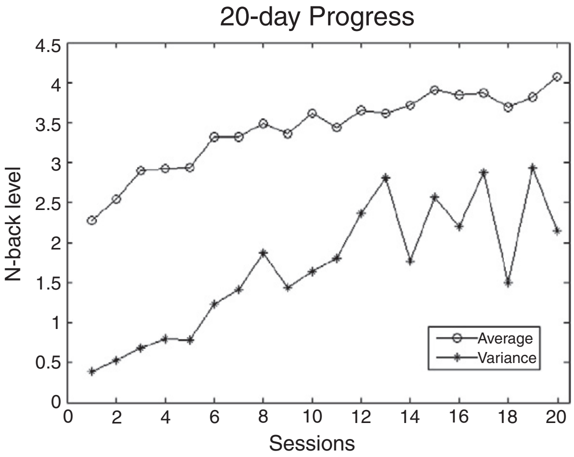


Fig. 3. Average performance on the training task and variance among participants in the 20-day condition.

Table 4
Effect sizes (Cohen's *d*) of test score difference.

Measure	8 day training (N=9)	20 day training (N=13)	8 day active control (N=15)	20 day active control (N=11)	8 day passive control (N=22)	20 day passive control (N=23)
Mill-Hill	+0.27	−0.25	−0.52	−0.04	−0.15	+0.08
Vocab	−0.72	−1.56	−1.19	−2.28	−1.42	−1.08
WBE	−0.15	−0.94	−0.16	−0.54	−0.57	−0.60
CPST	−0.62	−1.56	−1.28	−1.53	−1.34	−1.44
Identical Picture	−1.13	−0.64	−0.53	−1.28	−0.74	−0.61
Finding A's	−0.69	−0.22	+0.09	−0.12	−0.94	−0.90
Paper Folding	+0.42	+0.37	−0.25	+0.42	−0.12	+0.06
Card Rotation	+0.96	−0.06	+0.31	+0.40	+0.09	+0.40
Mental Rotation	−0.65	+0.08	−0.53	−0.11	−0.03	−0.41
OSPAN	+0.25	+0.15	+0.03	−0.06	+0.01	−0.02
RAPM	−0.05	−0.06	−0.00	−0.40	−0.47	−0.11

Positive values indicate higher post-test scores. WBE = Word Beginning and Ending; CPST = Colorado Perceptual Speed Test; OSPAN = Operation Span; RAPM = Raven's Advanced Progressive Matrices.

in the 8-day condition had approximately 34% improvement and those in the 20-day condition improved by 44%. The numbers suggested that participants from the current study and the original Jaeggi et al. (2008) study showed very similar performance on the training task; however, upon closer inspection of the data reported by Jaeggi et al., 2008, the authors collapsed post-test scores for all training groups and concluded a significant improvement in performance on intelligence tests after training based on an increase in about 2.5 points. This is misleading and inappropriate since not all participants took the same test for the purposes of detecting transfer effects. A comparison of improvement in the training task with participants in the replication study conducted in Taiwan (Jaeggi et al., 2010) was not included because the researchers modified the training task to accommodate the Chinese-speaking participants. The auditory modality of the dual n-back task was modified to syllables in the Mandarin phonetic system and instead of 20 blocks in each training session (~30 min), there were only 15 blocks which took about 17–20 min to complete (Jaeggi et al., 2010). Although our data on training improvement cannot be directly compared with those obtained from the two studies by Jaeggi et al. (2008, 2010), we argue that participants who trained their working memory in the current study improved on the training task just as much and that participants in the current study were just as motivated and committed as participants in the original study conducted by Jaeggi et al. (2008).

Participants in the current study and those in the studies conducted by Jaeggi et al. (2008, 2010) took the test under different administration procedures. RAPM was administered with no time constraint in the current study as recommended by the test provider, so participants were allowed to solve as many items as they could under no time pressure. Jaeggi and her colleagues administered their transfer tasks, the RAPM and BOMAT, with a time constraint—participants in their studies only had 10 min to solve as many items as they could (Jaeggi et al., 2008). In their first study, those in the 19-day training group answered about 4 more items on the BOMAT correctly at post-test (Jaeggi et al., 2008) and in their second study, the 20-day training group correctly answered 3 additional items in 16 min at post-test (Jaeggi et al., 2010). In their replication study, participants answered 2 additional items on the RAPM in 11 min after training for 20 days (Jaeggi et al., 2010). There was inconsistent usage of transfer tasks in the original study, where the researchers used the RAPM in the 8-day condition and not in the other training conditions. Participants who trained for 8-days showed no significant improvement on the RAPM at post-test (Jaeggi et al., 2008).

Many participants in the current study on average only needed about 10 min to attempt the items on the pre- and post-tests. This may be due to the population characteristic of the sample, which consisted of college students with high cognitive abilities. The pre- and post-tests in the current study were conducted in an hour. At pre-test, after the administration of the working memory task OSPAN and timed tests, participants had between 15 and 20 min before the end of the session to complete the Mill-Hill Vocabulary Test and RAPM—both of these tests were untimed. Participants were told that there were no time constraints and they could take as much time as they wanted to complete the items on both tests, so there were participants who took more than 20 min to complete both tests. Similarly, participants were given 15–20 min at the beginning of post-test session to work on the Mill-Hill and RAPM before the timed tests and OSPAN were administered. In essence, participants in the current study had as much time as those in the studies carried out by Jaeggi et al. (2008, 2010) with the added advantage of no time pressure exerted on the participants. Though Jaeggi et al. argued that the timed administration of RAPM/BOMAT in their studies was not an issue, the untimed administration of the same test in our study showed no significant improvements in RAPM scores.

The current study was designed to replicate and extend the original study by Jaeggi et al. (2008); thus, it was designed not only to detect an increase in scores but also to determine how the increase in performance arose should there be any, whether through improvements in verbal, perceptual or spatial rotation abilities following Johnson and Bouchard's (2005a) VPR model of intelligence. If general intelligence was improved after working memory training, it is imperative to know what underlying ability(ies) specifically led to an increase in general intelligence. The series of short, timed mental abilities tests administered in the current study were to provide additional information should there be an increase in the transfer task, RAPM. These tests were selected based on Johnson and Bouchard's (2005a) proposed model of intelligence, and exploratory factor analysis conducted on the test variables at pre-test (N=117) in the current study supported the model

(Table 5). However, results from the current study suggested no improvement overall in each of the three abilities.

Since the timed tests were employed to see if transfer effects could be detected due to changes in abilities such as perceptual speed and spatial ability, cutting the administration time limit to half was not considered a shortcoming seeing that it was necessary to accommodate for the (high) narrow range of participants in our study and to avoid ceiling effects. Participants in the current study had the chance to attempt an adequate number of items to provide enough variability despite the extremely short administration time. For instance, the paper folding test in the ETS allowed 3 min to complete each part (10 items), but participants in the current study were given 1.5 min to complete each part. Table 2 showed that the mean score at pre-test was about 5 (50% of the maximum score). There were individuals who could attempt almost all items within the time limit. The 30-second time limit for the Colorado Perceptual Speed Test, of which the recommended time limit was 60 s, only required participants to correctly identify the exact copy of the given stimulus, a measure of perceptual speed. Hence, any improvements detected in these timed tests would suggest an improvement in processing speed.

A major limitation of the study was the small sample size and possibly sample characteristic, which may have lowered the power of analyses conducted. When Jaeggi et al. (2010) repeated the study with 25 students who trained on the Raven's Advanced Progressive Matrices (RAPM) for 20 days, they obtained an effect size (Cohen's *d*) of 0.98. Additionally, participants in the Jaeggi et al. (2010) study were culturally different from the participants in the current study. Participants from the former study were undergraduates from a university in Taiwan (mean age = 19.4), while those from the current study were mostly American students attending a Midwestern university. The current study was designed according to the claims put forth by Jaeggi et al. (2008) as a study of replication and extension. In that study, participants were healthy, young adults who were slightly older (mean age = 25.6 years) than the current sample (mean age = 20.0), and they were recruited from a university in Bern, Switzerland. Effect sizes obtained from our study for RAPM were not as high as reported by Jaeggi et al. (2008, 2010)—$d = 0.65$ and $d = 0.98$ respectively. With such large effect sizes, the analysis of paired t-test could achieve a power of 0.80 with 10–12 participants. Referring to Table 4, the highest RAPM effect size ($d = 0.50$) was from the 8-day passive control group that had 22 participants and this achieved a power of 0.83. The 20-day training group (n = 13) had an effect size of 0.06 in RAPM, and to achieve a power of 0.80 this group would need more than 1700 participants. On the other hand, the effect size from the 20-day active control group with 11 participants was 0.40, and power could be improved by increasing the number of participants to 34. These observations led us to believe that the lack of improvements in the test variables was probably due to a combination of low sample size and differences in sample characteristics, of which participants in our study had restriction of range in intellectual ability.

Table 5
Exploratory factor analysis on pre-test variables.

Tests	Factor		
	1	2	3
MillHill1	**.917**	−.250	
Vocab1	**.660**	.181	
WordBE1	**.268**	.146	
IdentPix1	−.120	**.894**	.135
CPST1		**.556**	
FindA1	.138	**.343**	−.114
MentalRot1			**.763**
CardRot1			**.511**
PaperFold1			**.433**

Factor 1 corresponds to the Verbal factor, Factor 2 Perceptual factor and Factor 3 Mental Rotation factor in Johnson and Bouchard's (2005a, 2005b) VPR model. Values in bold refer to test variables that appropriately load on corresponding factors (verbal, perceptual or mental rotation).

Another limitation was that participants were given the option to switch groups if they knew they could not commit to the lab component. Though at pre-test participants were randomly assigned to groups, group assignment was not random at post-test due to the option of switching groups. Results could have been biased based on the "opt-out" option because we speculated that only the less motivated participants would have opted-out, and they may be the very individuals who could have benefited from the training and led to significant transfer effects. Referring to Table 1, the average N for each group originally was about 15. After group-switching from some participants, groups that involved the lab component had at least 10 participants except for the 8-day training group.

Limitations also included the group administered OSPAN, which may not have good discriminating validity because every participant is given the same amount of time (8 s) to evaluate the math equation and remember the word presented. Participants who were better and quicker at math would have had more time to remember and rehearse the words presented in the list, so these participants would probably have scored higher on the OSPAN. The unconventional administration of the task may have led to the failure of detecting WM capacity improvement after training.

More papers have emerged recently that supported the conclusions drawn from the current study. Redick et al. (2012) showed no evidence of improved intelligence after working memory training in a randomized, placebo-controlled study. The authors replicated the methods by Jaeggi et al. (2008) as much as they could while adding a series of tests (measures of fluid intelligence, crystallized intelligence, multi-tasking, working memory capacity and perceptual speed) to detect transfer effects. One of the strengths of the study was that they had good statistical power, yet they still failed to show transfer of training to other cognitive abilities that reflected *g* well. A meta-analysis by Melby-Lervåg and Hulme (2012) suggested that working memory training failed to generalize to other cognitive abilities. Colom et al. (2010) provided support for the stability of general intelligence using factor analyses that revealed no significant changes in the most *g*-loaded tests after training on either memory (short term memory and working memory) tasks or processing speed and attention tasks. In a large scale study (N = 11,430) by Owen et al. (2010) to test the validity of brain training, the researchers observed improvements in all the cognitive tasks that were trained but no transfer effects to untrained tasks, even when the untrained tasks shared similar cognitive functioning with those that were trained.

In conclusion, our results failed to show that intelligence can be improved simply by working memory training. Although sample size was a major limitation and there may be a significant difference in sample characteristic, of which participants in our study had restriction of range in intellectual ability, our results did not support the notion that intelligence in young, healthy adults can be improved simply by working memory training.

Role of the funding source

This work was supported by the Ministry of Higher Education, Malaysia as part of the corresponding author's dissertation. Besides financial contribution, the sponsor was not involved in any part of the preparation and execution of this work.

Acknowledgments

The authors would like to thank the following individuals for their support and contribution to the study: Professor Douglas Detterman, Dr. TJ McCallum, Professor H. Gerry Taylor, Michelle Blair, London Holt, Jasmine Core and Linda Lou.

References

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin*, *102*(1), 3–27.

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.

Basak, C., Boot, W. R., Voss, M. W., & Kramer, A. F. (2008). Can training in a real-time strategy video game attenuate cognitive decline in older adults? *Psychology and Aging*, *23*(4), 765–777.

Belmont, J. M., & Butterfield, E. C. (1977). The instructional approach to developmental cognitive research. In R. Kail, & J. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 437–481). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bergman-Nutley, S., Söderqvist, S., Bryde, S., Thorell, L. B., Humphreys, K., & Klingberg, T. (2011). Gains in fluid intelligence after training non-verbal reasoning in 4-year-old children: A controlled, randomized study. *Developmental Science*, *14*(3), 591–601.

Borella, E., Carretti, B., Riboldi, F., & De Beni, R. (2010). Working memory training in older adults: Evidence of transfer and maintenance effects. *Psychology and Aging*, *25*(4), 767–778.

Buschkuehl, M., Jaeggi, S. M., Hutchinson, S., Perrig-Chiello, P., Dapp, C., Muller, M., et al. (2008). Impact of working memory training on memory performance in old-old adults. *Psychology and Aging*, *23*(4), 743–753.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, *97*(3), 404–431.

Carretti, B., Borella, E., & De Beni, R. (2007). Does strategic memory training improve the working memory performance of younger and older adults? *Experimental Psychology*, *54*(4), 311–320.

Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports g and 10 broad factor. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5–21). Amsterdam: Pergamon.

Cattell, R. B. (1987). *Intelligence: Its structure, growth, and action.* New York: Elsevier.

Colom, R., Flores-Mendoza, C., & Rebollo, I. (2003). Working memory and intelligence. *Personality and Individual Differences*, *34*, 33–39.

Colom, R., Quiroga, M. A., Shih, P. C., Martinez, K., Burgaleta, M., Martinez-Molina, A., et al. (2010). Improvement in working memory is not related to increased intelligence scores. *Intelligence*, *38*, 497–505.

Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Science*, *7*(12), 547–552.

Deary, I. J. (2001). Human intelligence differences: Towards a recent history. *Trends in Cognitive Science*, *5*(3), 127–130.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, *333*, 959–964.

Ferrara, R. A., Brown, A. L., & Campione, J. C. (1986). Children's learning and transfer of inductive reasoning rules: Studies of proximal development. *Child Development*, *57*, 1087–1099.

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science*, *7*, 237–241.

Fry, A. F., & Hale, S. (2000). Relationships among processing speed, working memory, and fluid intelligence in children. *Biological Psychology*, *54*, 1–34.

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

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Science*, *11*(6), 236–242.

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*, F9–F15.

Holmes, J., Gathercole, S. E., Place, M., Dunning, D. L., Hilton, K. A., & Elliott, J. G. (2010). Working memory deficits can be overcome: Impacts of training and medication on working memory in children with ADHD. *Applied Cognitive Psychology*, *24*(6), 827–836.

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., Studer-Luethi, B., Buschkuehl, M., Su, Y. F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning—Implications for training and transfer. *Intelligence*, *38*, 625–635.

Johnson, W., & Bouchard, T. J. (2005a). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, *33*, 393–416.

Johnson, W., & Bouchard, T. J. (2005b). Constructive replication of the visual-perception-image rotation model in Thurstone's (1941) battery of 60 tests of mental ability. *Intelligence*, *33*, 417–430.

Johnson, W., te Nijenhuis, J., & Bouchard, T. J. (2007). Replication of the hierarchical visual-perceptual-image rotation model in de Wolff and Buiten's (1963) battery of 46 tests of mental ability. *Intelligence*, *35*, 69–81.

Jurden, F. H. (1995). Individual differences in working memory and complex cognition. *Journal of Educational Psychology*, *87*, 93–102.

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

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 66–71.

Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science*, *12*(6), 978–990.

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlstrom, K., et al. (2005). Computerized training of working memory training in children with ADHD—A randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, *44*(2), 177–186.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology*, *24*(6), 781–791.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*, 389–433.

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

Mackey, A. P., Hill, S. S., Stone, S. I., & Bunge, S. A. (2011). Differential effects of reasoning and speed training in children. *Developmental Science*, *14*(3), 582–590.

Melby-Lervåg, M., & Hulme, C. (2012). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, http://dx.doi.org/10.1037/a0028228 (Advance online publication).

Minear, M., & Shah, P. (2008). Training and transfer effects in task switching. *Memory and Cognition*, *36*(8), 1470–1483.

Moody, D. E. (2009). Can intelligence be increased by training on a task of working memory? *Intelligence*, *37*, 327–328.

Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review*, *18*, 46–60.

Oberauer, K., Schulze, R., Wilhelm, O., & Süβ, H. -M. (2005). Working memory and intelligence—Their correlation and their relation: Comment on Ackerman, Beier and Boyle (2005). *Psychological Bulletin*, *131*, 61–65.

Orrell, M., & Sahakian, B. (1995). Education and dementia: Research evidence supports the concept of “use it or lose it”. *British Medical Journal*, *310*, 951–952.

Owen, A. M., Hampsire, A., Grahn, J. A., Stenton, J., Dajani, S., Burns, A. S., et al. (2010). Putting brain training to the test. *Nature*, *465*(7299), 775–778.

Raven, J. C. (1990). *Advanced progressive matrices: Sets I, II.* Oxford: Oxford Univ Press.

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., et al. (2012). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology. General*, http://dx.doi.org/10.1037/a0029082 (Advance online publication).

Rueda, M. R., Rothbabrt, M. K., McCandliss, B. D., Saccomanno, L., & Posner, M. I. (2005). Training, maturation, and genetic influences on the development of executive attention. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(41), 14931–14936.

Salthouse, T. A. (1993). Influence of working memory on adult age differences in matrix reasoning. *British Journal of Psychology*, *84*, 171–199.

Schmiedek, F., Lovden, 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*(27), 1–10.

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.

Shepard, R., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701–703.

Spitz, H. (1986). *The raising of intelligence: A selected history of attempts to raise retarded intelligence.* Hillsdale, NJ: Erlbaum.

St. Clair-Thompson, H., Stevens, R., Hunt, A., & Bolder, E. (2010). Improving children's working memory and classroom performance. *Educational Psychology*, *30*(2), 203–219.

Stauffer, J. M., Ree, M. J., & Carreta, T. R. (1996). Cognitive-components tests are not much more than *g*: An extension of Kyllonen's analyses. *The Journal of General Psychology*, *123*, 193–205.

Thorell, L. B., Lindqvist, S., Bergman-Nutley, S., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science*, *12*(1), 106–113.

Tranter, L. J., & Koutstaal, W. (2008). Age and flexible thinking: An experimental demonstration of the beneficial effects of increased cognitively stimulating activity on fluid intelligence in healthy older adults. *Aging, Neuropsychology and Cognition*, *15*, 184–207.

Tucker, P., & Warr, P. (1996). Intelligence, elementary cognitive components, and cognitive styles as predictors of complex performance. *Personality and Individual Differences*, *21*, 91–102.

Unsworth, N., & Engle, R. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence*, *33*, 67–81.

Van der Molen, M. J., Van Luit, J. E. H., Van der Molen, M. W., Klugkist, I., & Jongmans, M. J. (2010). Effectiveness of a computerized working memory training in adolescents with mild to borderline intellectual disabilities. *Journal of Intellectual Disability Research*, *54*, 433–447.

Verguts, T., & De Boeck, P. (2002). On the correlation between working memory capacity and performance on intelligence tests. *Learning and Individual Differences*, *13*, 37–55.

Wahlsten, D. (1997). The malleability of intelligence is not constrained by heritability. In B. Devlin, S. E. Fienberg, & K. Roeder (Eds.), *Intelligence, genes, and success: Scientists respond to The Bell Curve* (pp. 71–87). New York: Springer.