



SHORT REPORT

Does working memory training lead to generalized improvements in children with low working memory? A randomized controlled trial

Darren L. Dunning,¹ Joni Holmes² and Susan E. Gathercole²

1. Department of Psychology, University of York, UK

2. MRC Cognition and Brain Sciences Unit, Cambridge, UK

Abstract

Children with low working memory typically make poor educational progress, and it has been speculated that difficulties in meeting the heavy working memory demands of the classroom may be a contributory factor. Intensive working memory training has been shown to boost performance on untrained memory tasks in a variety of populations. This first randomized controlled trial with low working memory children investigated whether the benefits of training extend beyond standard working memory tasks to other more complex activities typical of the classroom in which working memory plays a role, as well as to other cognitive skills and developing academic abilities. Children aged 7–9 years received either adaptive working memory training, non-adaptive working memory training with low memory loads, or no training. Adaptive training was associated with selective improvements in multiple untrained tests of working memory, with no evidence of changes in classroom analogues of activities that tax working memory, or any other cognitive assessments. Gains in verbal working memory were sustained one year after training. Thus the benefits of working memory training delivered in this way may not extend beyond structured working memory tasks.

Introduction

Working memory (WM) is the cognitive system responsible for the temporary storage of information required to support ongoing everyday activities such as following instructions, mental arithmetic, and problem-solving (Adams & Hitch, 1997; Gathercole, Durling, Evans, Jeffcock & Stone, 2008b; Oberauer, Suß, Wilhelm & Wittman, 2008). Children with poor WM skills at school entry are at high risk of educational underachievement in reading and mathematics (Gathercole, Brown & Pickering, 2003), and typically make poor progress in all assessed areas of the academic curriculum (Geary, 2004; Swanson & Sachse-Lee, 2001). These difficulties in learning may be unsurprising due to the heavy WM demands of the classroom. Children with poor WM frequently fail to complete common classroom activities

that require large amounts of information to be held in mind, experience difficulties following multi-step instructions, and have problems keeping their place in demanding and complex activities (Engle, Carullo & Collins, 1991; Gathercole & Alloway, 2008; Gathercole, Alloway, Kirkwood, Elliott, Holmes & Hilton, 2008a).

Early interventions that could overcome these relatively common cognitive and learning difficulties would clearly be of great potential benefit to these children, and there is accumulating evidence that intensive WM training can substantially boost performance on non-trained WM tasks (e.g. Klingberg, Fernell, Olesen, Johnson, Gustafsson, Dahlstrom, Gillberg, Forssberg & Westerberg, 2005). Training programmes typically require individuals to train intensively for a continued period on tasks adapted to their current performance. Cogmed Working Memory Training (CWMT;

Address for correspondence: Darren L. Dunning, Department of Psychology, University of York, York YO10 5DD, UK; e-mail: darren.dunning@york.ac.uk

Klingberg, Forssberg & Westerberg, 2002, Klingberg *et al.*, 2005) is the most widely used of these, and has been linked with enhancements in WM in children with ADHD (Klingberg *et al.*, 2005; Holmes, Gathercole, Place, Dunning, Hilton & Elliot, 2010b), special educational needs in reading (Dahlin, 2010), and attentional problems (Mezzacappa & Buckner, 2010). There is also preliminary evidence that it is beneficial for children with low WM who have no other recognized learning difficulties (Holmes, Gathercole & Dunning, 2009).

The methodologies adopted in research studies of WM training have recently been the focus of debate (Gathercole, Dunning & Holmes, 2012; Klingberg, 2012; Melby-Lervåg & Hulme, 2013; Shipstead, Redick & Engle, 2010, 2012). Shortcomings levelled at the field include failing to employ outcome measures with high construct validity and not using multiple measures of each construct, not randomly allocating participants to condition, inadequate sample sizes and either the lack of a control group or failing to include both a no treatment baseline condition and an active comparison intervention to control for non-specific factors. The present study is the first randomized controlled trial (RCT) of CWMT with children with low WM that overcomes these methodological shortcomings. Both active and passive control interventions were included, schools were randomly allocated to condition, and multiple and well-validated measures of WM, reading, maths, and other cognitive skills were administered by researchers who were blind to intervention condition.

We investigated whether two fundamental aspects of transfer of training would withstand the challenges of this rigorous methodology: transfer to other conventional WM tasks, and transfer to analogues of WM-loaded classroom activities in which low WM children are known to be impaired (Gathercole, Lamont & Alloway, 2006, Gathercole *et al.*, 2008b). The latter of these was measured by three tasks. In one, the child counted the number of words in a spoken sentence before attempting verbatim recall. This activity involved the temporary storage of the sentence while the words were being counted. The second task required the child to identify a pair of rhyming words in short spoken poems. Here the child must monitor and compare the phonological forms of a potentially large number of word pairs, hence requiring off-line support from WM. In the third task, the child attempted to carry out sequences of spoken instructions using familiar classroom items. Children with low WM have been shown to improve on this task following CWMT (Holmes *et al.*, 2009).

We also examined whether selective cognitive enhancements with adaptive training extend to other cognitive functions including speed of visual search, sustained

attention, IQ, as well as mathematics and reading. Whilst improvements in WM tasks that closely resemble the trained activities are reported consistently (e.g. Klingberg *et al.*, 2005; Holmes *et al.*, 2009), the evidence for transfer to tasks that share little overlap with the structure and content of trained activities while drawing on hypothesized common processes is mixed (e.g. Jaeggi, Buschkuhl, Jonides & Shah, 2011). Evaluation of the impact of training on children's learning is vital as it is central to the practical value of training. A small number of studies have reported changes in reading and mathematics performance after training, but in each case an RCT methodology was not employed. For example, Holmes *et al.* (2009) reported improvements in maths scores 6 months after adaptive training but did not conduct a similar follow-up assessment for the control group, raising the possibility that gains were a case of repeat testing. The improvements reported in reading comprehension by Dahlin (2010) were significant in comparison to a no-contact rather than an active control group, which guards against test-retest but not the Hawthorne effect (see McCarney, Warner, Illife, van Haselen, Griffin & Fisher, 2007). The present study provided the first substantial test of whether training in low WM children leads to improvements in learning using measures widely claimed to depend on WM such as mathematical reasoning (Swanson & Sacshe-Lee, 2001) and reading comprehension (e.g. Cain, Oakhill & Bryant, 2004).

The sustainability of any training gains and the potential impact of training on learning progress in literacy and in mathematics were assessed 12 months after training. Previous non-RCT studies have shown that the positive effects of training remain intact after 6 months (e.g. Holmes *et al.*, 2009). The question here is whether such gains are sustained over a longer period using a more robust methodology.

Method

Participants and procedure

To identify children with low WM, 810 children (425 boys, age $M = 8y\ 5m$, $SD = 7.91m$) attending nine schools in the North-East of England were included in the screening phase. In these schools, 29% received free school meals (nationally, 19%), and 11% had special educational needs (nationally, 21%). All children were screened on two tests of the Automated Working Memory Assessment (AWMA; Alloway, 2007): a verbal test, backward digit recall, in which the children attempted to recall spoken strings of digits in reverse

order, and a visuo-spatial test, Mr X, which involved the recall of a series of locations, interspersed with mental rotation decisions.

Ninety-four children with English as their first language (47 boys, age $M = 8y 5 m$, $SD = 7.97 m$) with standard scores at or below the 15th centile on both screening measures were identified as having low WM. These children had normal or corrected-to-normal vision and hearing. No other exclusion criteria were applied. Written consent was obtained both from the participating schools and the parents/guardians of the children.

These children completed pre-training assessments (T1) and were assigned to one of three groups: adaptive training, non-adaptive training or no intervention. Participants in the adaptive and non-adaptive groups then completed 6 weeks' training. Post-training assessments were completed after training (T2) for the adaptive and non-adaptive groups, and 6 weeks after pre-assessment for the no intervention group.

On the basis of schools' willingness to participate, 15 children in the adaptive group (eight boys, mean age 9y 3 m, $SD 4.97$) and 19 in the non-adaptive group (eight boys, mean age 9y 6 m, $SD 7.15$) were re-tested 12 months after training (T3). A subset of measures was used to reduce testing time.

All assessments were conducted by research assistants blind to the intervention status of the children. Statistical power was .94 with a p level of .05 (Erdfelder, Faul & Buchner, 1996) for the main study, and .62 for the 12-month follow-up.

Participating schools were randomly assigned to adaptive training, non-adaptive training or no intervention conditions. This cluster randomization procedure minimizes contamination between the two training groups and avoids the dilution of effects that would have occurred with random assignment of pupils to conditions within schools (Torgerson & Torgerson, 2008). Although necessary given co-training of children from both training groups in school, cluster randomization is associated with decreased variability of individuals within clusters due to shared factors other than condition, such as having the same teacher. Intra-cluster correlations (ICC) were calculated for T1 measures (see Table 1) to measure the variance between clusters, and consequently determine how similar the outcome would be had individual randomization been used (Koch, 1982). ICCs of .5–.6 indicate moderate agreement between clusters, .7–.8 show a strong agreement, and ICCs of over .8 almost perfect agreement (Landis & Koch, 1977). The ICCs in this study reflect an acceptably low level of risk that clustering would generate spurious differences between conditions (Kerry & Bland, 1998).

Table 1 T1 Intra-cluster correlations

	Between cluster variance (χ^2)	Within cluster variance (χ^2)	Correlation coefficient
Memory measures			
AWMA			
Verbal STM ¹	381.95	565.02	.68
Visuo-spatial STM ¹	278.55	378.25	.74
Verbal WM ¹	77.44	121.54	.64
Visuo-spatial WM ¹	45.50	90.46	.50
Classroom-based tasks			
Following instructions ²	30.91	45.78	.68
Rhyme recall ²	6.62	10.77	.62
Sentence counting (processing) ²	31.90	44.20	.72
Sentence counting (storage) ²	11.14	21.27	.52
Ability			
Intelligence			
Verbal IQ ¹	464.25	665.20	.70
Performance IQ ¹	315.38	434.92	.73
Maths			
Number operations ¹	681.56	848.87	.80
Maths reasoning ¹	320.91	512.46	.63
Reading			
Basic reading ¹	506.07	700.14	.72
Reading accuracy ¹	351.80	526.24	.67
Reading comprehension ¹	437.54	603.31	.73
Reading rate ¹	272.58	494.74	.55
Written expression ³	465.14	597.75	.78
Visual scanning and attention			
Visual scanning			
Items correct ⁴	1.03	1.46	.70
Sustained attention			
Omissions ⁵	989.34	1473.73	.67
Commissions ⁵	152.77	213.78	.72
Hit Rate ⁵	325.53	517.23	.63

¹Standard score. ²Trials correct. ³Raw scores. ⁴Seconds per correct response. ⁵ t -scores.

Working memory training

The adaptive training group received CWMT for 20–25 sessions. Each session lasted for between 30 and 45 minutes and consisted of training on eight exercises, with 15 trials on each, giving a total of 120 trials to be completed in each training session. The eight training exercises were selected daily from a bank of 12. CWMT is a commercially available product and the exercises presented in each training session were delivered in a preset manner. Task difficulty was adjusted on a trial-by-trial basis to match the participant's current WM span. A high-score chart allowed participants to gauge their performance level. For full details see www.cogmed.com/rm.

Children in the non-adaptive group were trained on a version of CWMT developed for trial evaluations (Klingberg *et al.*, 2005). This version was identical to the adaptive version of the programme apart from training tasks were set at a low span level of two throughout the training period, and did not increase in difficulty with task success and, as the difficulty level was fixed, there was no high-score chart.

Training was conducted in school with groups of 6–12 individuals, under researcher supervision. Children in both training groups received small rewards such as stationery items for every five training sessions completed. Inevitably, all motivational features could not be equated across the two training programmes. The percentage of total time spent on task recorded on the Cogmed log (relative to total training time) was comparable across groups (adaptive, 80%; non-adaptive, 82%).

Measures

Working memory

At T1 eight standardized subtests from the AWMA were administered to each child over two testing sessions – two tests each of verbal STM (digit and word recall), visuo-spatial STM (dot matrix and block recall), verbal WM (counting recall and backward digit recall) and visuo-spatial WM (Mr X and spatial span). At T2, four tasks were repeated from T1 and four were introduced for the first time: verbal STM (digit and nonword recall), visuo-spatial STM (dot matrix and mazes memory), verbal WM (listening recall and backward digit recall) and visuo-spatial WM (Mr X and odd-one-out). Correlations between the four pairs of non-repeated subtests tapping common constructs were: word recall–nonword recall = .60, block recall–mazes memory = .49, counting recall–listening recall = .53, spatial span–odd-one-out = .59 (Alloway, Gathercole & Pickering, 2006). Composite scores were derived for each of the four aspects of WM by calculating the mean standard scores from the AWMA for each pair to increase the robustness of the assessments.

Classroom-based tasks

Three tests based on classroom activities known to vary with WM skill employed by Gathercole *et al.* (2006, 2008b) were administered at each testing point.

Following Instructions. An array of 15 coloured objects (folders, boxes, pencils, rulers and erasers that were coloured red, blue and yellow) were placed in front of the participant. On each trial, the experimenter read aloud a

set of instructions relating to a subset of the objects, which participants attempted to follow. A span method was used, with each span consisting of a block of six trials. Testing started at one action (e.g. *touch the red pencil*) and increased by one action per block until the participant was unable to complete four trials in a block correctly. The total number of trials correct was calculated.

Detecting rhymes. On each trial, the experimenter read aloud two-line poems containing between 4 and 10 words (mean = 6.57) such as *I like cakes, my mum bakes*. Participants were asked to recall the two rhyming words in the poem. There were seven trials and the total number of pairs of rhyming words correctly recalled was calculated. The difficulty level of this task did not vary.

Sentence counting and recall. On each trial, a sentence containing between two and nine words (mean = 5.71) was spoken aloud (e.g. *pigs like to roll around in the mud*) and participants were asked to count the number of words in the sentence, say the total aloud, and then repeat the sentence verbatim. Fourteen sentences were presented.

Ability tests

At T1 and T2, participants completed the Wechsler Abbreviated Scales of Intelligence (WASI; Wechsler, 1999), which consists of four subtests: Similarities and Vocabulary (verbal IQ), and Matrix Reasoning and Block Design (performance IQ). They also completed the Mathematical Reasoning and Number Operations subtests of the Wechsler Objective Number Dimensions test (WOND; Wechsler, 1996) and the Basic Reading subtest of the Wechsler Objective Reading Dimensions (WORD; Wechsler, 1993). The Neale Analysis of Reading Ability test (NARA; Neale, 1997) yielded measures of reading accuracy, reading comprehension and reading rate. The Written Expression subtest from the Kaufman Test of Educational Attainment (KTEA; Kaufman & Kaufman, 2004) was also administered. At T3, participants completed the Similarities and Matrix Reasoning subtests of the WASI, the Basic Reading subtest of the WORD, the Number Operation subtest of the WOND, and the NARA.

Other cognitive assessments

The Continuous Performance Test (CPT; Conners & Multi-Health Systems Staff, 2004), a computerized assessment in which children respond to target letters, provided a measure of sustained attention; omissions,

commissions and hit rate were recorded. Visual scanning speed was assessed using the Visual Scanning subtest of the Delis-Kaplan Executive Function System (D-KEFS; Delis, Kaplan & Kramer, 2001). This involved the detection of visual targets in a search task. Completion time (in seconds) and the number of errors were used to calculate the time taken per correct target. Both tasks were administered at T1, T2 and T3.

Results

Table 2 shows descriptive statistics for T1 and T2 scores by group. Performance at T1, T2 and T3 is shown in Table 4 for the subset of children from the adaptive and non-adaptive groups who were followed up one year after training.

One-way ANOVAs were performed for each measure at T1 as a function of group (adaptive, non-adaptive and no intervention). Significant group differences were found for visuo-spatial STM, $F(2, 91) = 4.03$, $p = .02$, with the non-adaptive group outperforming the no-intervention group, $F(1, 58) = 7.82$, $p = .01$, and for verbal WM, $F(2, 91) = 4.20$, $p = .02$, with the non-adaptive, $F(1, 62) = 6.91$, $p = .01$, and no intervention groups, $F(1, 62) = 6.52$, $p = .01$, scoring higher than the adaptive group. There were also significant group differences at T1 for commissions on the CPT, $F(2, 88) = 5.31$, $p = .01$, with the no intervention group significantly higher than the adaptive group, $F(1, 59) = 8.61$, $p = .01$, and also for the Visual Scanning task, $F(2, 76) = 3.49$, $p = .04$, with the adaptive group scoring significantly lower than the non-adaptive, $F(1, 61) = 5.04$, $p = .03$, and no intervention groups, $F(1, 62) = 6.03$, $p = .02$.

To test group effects on training gains, general linear models were performed separately for the different T2 measures with scores at T2 entered as the dependent variable and scores at T1 and group as independent variables. For the measures with significant group differences at baseline, group*T1 interaction terms were added to the model. Interaction terms were created by first centring T1 scores and multiplying the centred score by the group dummy variables. Thus, the independent variables entered for visuo-spatial STM, verbal WM, CPT commissions and processing speed (the measures on which there were group differences at T1) were T1 scores, group, the T1*group interactions, and the centred T1 score.

Table 3 summarizes the outcome of these analyses. Significant training effects were observed for visuo-spatial STM, verbal WM and visuo-spatial WM, with scores significantly greater for the adaptive than both

other groups at T2 (see Figure 1). This pattern of results was replicated when only the WM measures that were repeated at T1 and T2 were included. Again, scores were significantly greater for the adaptive group than both other groups for visuo-spatial STM, verbal WM and visuo-spatial WM: Verbal STM (Digit Recall), adaptive–non-adaptive, $p = .19$, adaptive–no intervention, $p = .64$; Visuo-spatial STM (Dot Matrix) adaptive–non-adaptive, $p < .01$, adaptive–no intervention, $p < .01$; Verbal WM (Backward Digit Recall) adaptive–non-adaptive, $p = .01$, adaptive–no intervention, $p < .01$; Visuo-spatial WM (Mr X) adaptive–non-adaptive, $p = .05$, adaptive–no intervention, $p < .01$. Basic Reading scores at T2 were also significantly predicted by group, with the no intervention group outperforming the other two groups. There were no other significant training effects.

Baseline performance for the subset of children from the adaptive and non-adaptive groups retested 12 months after training (T3) was analysed in series of one-way ANOVAs comparing scores for each measure at T1 that was re-administered at T3. There were no significant group differences at baseline. To test the effect of group on outcome, general linear models were performed separately for each of the T3 measures, with performance at T3 entered as the dependent variable and performance at T1 and group entered as independent variables. Group was a significant predictor of T3 scores for verbal WM and the processing aspect of the sentence counting task, with the adaptive group outperforming the non-adaptive group (see Table 4).

A Bonferroni correction reduces the probability criterion for significance to .003 at both T2 for the whole sample and T3 for those included in the follow-up. The only finding that withstands this correction is verbal WM at T2. However, effect sizes (Cohen's d) were substantial at T2, ranging from .67 to .99 between the adaptive and non-adaptive groups and .57 to 1.63 between the adaptive and no intervention groups.

Discussion

In this RCT, adaptive WM training significantly boosted performance on untrained WM tasks in children with low WM. This enhancement was substantial in magnitude and was partially sustained for 12 months. Children who completed adaptive training made significantly greater improvements in tests of visuo-spatial STM and verbal and visuo-spatial WM than either children who completed a non-adaptive version of training or those who received no intervention. There were few significant differences between the two control groups, suggesting that the gains associated with adaptive training are

Table 2 Pre- and post-training scores by group

	Adaptive training						Non-adaptive training						No intervention						
	Pre-assessment (T1)		Post-assessment (T2)		Pre-assessment (T1)		Post-assessment (T2)		Pre-assessment (T1)		Post-assessment (T2)		Pre-assessment (T1)		Post-assessment (T2)				
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	
Working memory																			
AWMA																			
Verbal STM ¹	34	89.04	15.05	34	93.56	11.40	30	85.20	14.71	30	88.22	12.06	30	84.27	12.38	30	91.72	10.07	
Visuo-spatial STM ¹	34	86.59	10.24	34	96.97	15.35	30	91.16	10.04	30	90.45	9.23	30	83.55	10.90	30	87.03	12.19	
Verbal WM ¹	34	81.41	5.53	34	96.25	11.08	30	85.48	6.85	30	90.47	9.23	30	85.58	7.49	30	85.47	9.22	
Visuo-spatial WM ¹	34	86.35	6.63	34	99.94	10.25	30	88.35	5.60	30	95.43	8.96	30	84.20	7.52	30	90.13	9.68	
Classroom -based tasks																			
Following instructions ²	33	15.42	4.18	33	17.94	5.06	28	15.82	4.20	28	15.00	4.42	17	15.76	4.05	17	15.94	3.82	
Rhyme recall ²	32	3.72	2.17	32	4.53	2.06	30	3.63	2.19	30	4.63	1.67	30	4.23	1.91	30	4.60	1.87	
Sentence counting (storage) ²	34	6.76	3.75	34	9.71	3.33	30	8.80	3.79	30	10.07	3.22	30	9.47	3.17	30	9.07	3.33	
Sentence counting (processing) ²	34	8.59	4.00	34	10.15	3.67	30	9.57	3.08	30	10.27	3.30	30	10.00	2.56	30	9.80	3.27	
Ability																			
IQ																			
Verbal IQ ¹	34	90.82	10.71	34	91.00	13.94	26	92.77	16.44	26	90.62	15.75	30	87.97	17.29	30	90.17	15.99	
Performance IQ ¹	34	92.41	13.44	34	91.68	13.48	27	88.89	8.47	27	92.22	11.27	30	87.07	11.83	30	90.13	10.87	
Maths																			
Number operations ¹	33	94.40	14.49	33	94.15	13.76	29	90.07	14.36	29	92.90	16.16	30	89.43	14.15	30	91.10	14.17	
Maths reasoning ¹	33	90.39	15.17	33	87.36	14.89	29	89.14	16.61	29	87.76	13.64	30	82.97	10.81	30	81.90	12.16	
Reading																			
Basic reading ¹	34	88.71	12.88	34	88.85	13.93	30	90.69	17.47	29	89.07	15.86	30	88.10	14.86	30	92.03	15.83	
Reading accuracy ¹	30	86.80	12.12	30	88.80	12.94	28	88.61	14.40	28	90.43	12.32	28	86.71	15.16	28	88.21	12.12	
Reading comprehension ¹	30	82.63	11.97	30	85.57	12.54	28	86.96	14.55	28	87.00	12.56	28	84.43	13.56	28	84.46	12.58	
Reading rate ¹	29	101.38	13.90	29	104.69	14.21	27	103.15	16.27	27	106.04	13.54	27	101.70	15.47	27	101.52	16.66	
Written expression ³	32	163.72	13.29	32	167.88	12.42	27	173.93	11.45	27	173.11	12.66	16	168.00	10.28	16	168.69	11.63	
Visual scanning and attention																			
Visual scanning																			
Items correct ⁴	33	2.10	.75	33	1.68	.58	30	1.74	.61	29	1.55	.48	30	1.71	.61	30	1.65	.52	
Attention																			
Omissions ⁵	33	68.41	26.20	33	62.26	15.03	27	66.58	16.41	29	64.047	14.77	22	77.50	25.01	22	80.756	32.71	
Commissions ⁵	33	47.45	9.86	33	50.92	8.65	27	51.67	6.38	29	54.44	8.61	22	54.63	6.56	22	53.38	7.44	
Hit Rate ⁵	33	63.52	13.93	33	58.46	11.70	27	60.57	14.67	29	59.32	13.91	22	69.31	13.62	22	65.76	15.19	

¹Standard score. ²Trials correct. ³Raw scores. ⁴Seconds. ⁵*t*-scores.

Table 3 T1–T2 Group comparisons

	Model		Group coefficients					
	R^2	F	Adaptive vs. non-adaptive		Adaptive vs. no intervention		Non-adaptive vs. no intervention	
			t	d	t	d	t	d
Working memory								
AWMA								
Verbal STM ¹	.48	27.56***	1.32	.14	.43	.28	1.72	.44
Visuo-spatial STM ¹	.31	10.16***	2.50*	.87	2.73**	.57	.187	.36
Verbal WM ¹	.37	12.96***	3.66***	.99	6.43***	1.63	2.83**	.59
Visuo-spatial WM ¹	.27	11.00***	2.38*	.67	3.89***	.77	1.38	.14
Classroom-based tasks								
Following instructions ²	.29	10.65***	2.59*	.71	1.21	.12	1.16	.98
Rhyme recall ²	.55	55.09***	.28	.11	1.50	.26	.59	.14
Sentence counting ²	.37	17.89***	.59	.32	1.85	.54	1.22	.31
Sentence counting (processing) ²	.49	28.80***	.93	.70	2.37***	1.10	1.44*	.61
Ability								
IQ								
Verbal IQ ¹	.60	44.70***	1.70	.27	.22	.16	1.45	.39
Performance IQ ¹	.62	49.99***	1.55	.50	1.15	.54	.43	.08
Maths								
Number operations ¹	.72	75.58***	1.03	.40	.34	.22	.69	.14
Maths reasoning ¹	.69	66.68***	.60	.20	.01	.24	.57	.04
Reading								
Basic reading ¹	.85	172.01***	1.52	.36	2.26*	.62	3.62***	.85
Reading accuracy ¹	.83	135.20***	.10	.03	.67	.09	.55	.05
Reading comprehension ¹	.72	71.55***	1.12	.40	1.55	.45	.40	.00
Reading rate ¹	.54	31.29***	.25	.04	1.16	.32	1.36	.30
Written expression ³	.73	62.44***	1.99*	.69	1.42	.57	.4	.23
Visual scanning and attention								
Visual scanning								
Items correct ⁴	.33	8.50***	.40	.40	1.28	.24	.84	.19
Attention								
Omissions ⁵	.23	7.80***	.47	.19	2.53*	.32	1.98*	.24
Commissions ⁵	.36	8.97***	.22	.09	1.05	.70	1.22	.59
Hit Rate ⁵	.44	21.02***	1.78	.31	1.39	.14	.29	.19

¹Standard score. ²Trials correct. ³Raw scores. ⁴Seconds. ⁵ t -scores. * $p < .05$; ** $p < .01$; *** $p < .001$.

unlikely to be due to expectancy effects. This is the first double-blinded RCT study on this key group of poor learners that meets stringent criteria for intervention research (e.g. Shipstead *et al.*, 2010), and the results reinforce outcomes of an earlier training study with children with low WM (Holmes *et al.*, 2009). Once again, WM training failed to enhance performance on tests of verbal STM (Holmes *et al.*, 2009; Holmes, Gathercole & Dunning, 2010a). This is consistent with evidence that unlike the other measures, verbal STM tasks tap a highly specialized component of WM that places minimal demands on the executive control of WM and is linked with vocabulary acquisition (Baddeley, Gathercole & Papagno, 1998), rather than more general academic learning.

The boost in verbal WM performance with adaptive training persisted for 12 months for the subgroup of children re-tested at this point, beyond the 6-month period previously investigated (e.g. Holmes *et al.*, 2009;

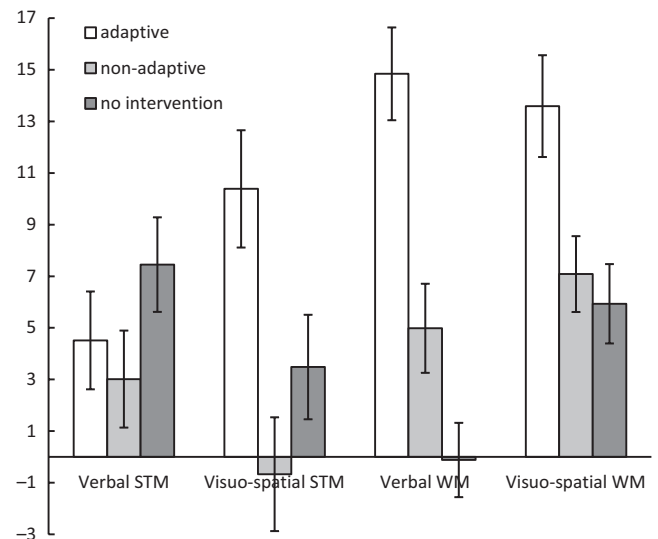


Figure 1 Changes in working memory standard scores T1–T2, by group, with standard errors.

Table 4 *T1, T2 and T3 scores and group coefficients for the children included in the 12 month follow-up*

	Adaptive training						Non-adaptive training						Group comparisons (T1-T3)												
	Pre-assessment (T1)			Post-assessment (T2)			12 month follow-up (T3)			Pre-assessment (T1)			Post-assessment (T2)			12 month follow-up (T3)			Group coefficients						
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>		<i>R</i> ²	<i>F</i>	<i>t</i>	<i>d</i>		
Memory measures																									
AWMA																									
Verbal STM (Digit recall) ¹	15	96.60	19.95	15	95.27	16.74	15	100.20	18.38	19	84.95	13.89	19	83.42	15.00	19	91.68	12.71	.58	21.27**	.88		.12		
Visuo-spatial STM (Dot matrix) ¹	15	88.13	10.97	15	113.00	27.50	15	101.60	17.81	19	91.68	13.65	19	96.53	14.01	19	95.63	14.91	.15	2.71	1.40		.55		
Verbal WM (Backward digit recall) ¹	15	74.87	7.80	15	101.40	14.04	15	98.33	9.70	19	76.37	9.54	19	91.84	12.56	19	88.21	10.22	.22	4.45*	2.95**		1.16		
Visuo-spatial WM (MrX) ¹	15	81.13	5.04	15	100.00	11.67	15	96.87	8.81	19	81.37	5.92	19	93.16	9.94	19	95.89	10.78	.00	.05	.27		.26		
Classroom-based tasks																									
Following instructions ²	15	16.13	4.67	14	19.14	5.72	14	18.07	4.03	19	16.58	4.65	17	15.29	4.66	17	16.47	3.36	.25	4.77*	1.14		.23		
Rhyme recall ²	15	4.07	2.67	15	5.20	1.86	13	5.31	1.84	19	4.00	1.94	19	4.89	1.79	16	5.19	1.87	.36	7.28**	.22		.07		
Sentence counting (storage) ²	15	9.67	3.85	15	10.73	3.37	13	11.61	2.60	19	9.79	2.72	19	10.74	3.57	16	11.81	2.14	.44	10.24**	.15		.01		
Sentence counting (processing) ²	15	8.07	3.54	15	10.87	3.14	13	11.92	2.33	19	9.79	3.31	19	11.11	2.56	16	11.06	3.11	.49	9.04**	2.39*		1.08		
Ability measures																									
IQ																									
Similarities ³	15	49.53	8.38	15	48.87	10.6	14	48.07	9.22	19	51.16	9.49	19	50.47	11.18	17	50.54	15.57	.20	3.58*	.35		.67		
Matrix Reasoning ³	15	46.13	10.18	15	45.20	10.23	14	45.07	9.79	19	42.32	7.98	19	45.79	9.85	17	42.47	7.29	.72	35.62**	.66		.53		
Maths																									
Maths reasoning ¹	15	97.67	15.44	15	95.87	12.34	15	96.87	10.49	19	93.32	12.55	19	97.16	13.60	16	93.25	10.76	.56	20.03**	.14		.27		
Reading																									
Basic reading ¹	15	93.20	16.88	15	93.53	16.98	14	88.79	14.41	19	96.89	16.36	19	94.37	13.95	17	92.88	14.30	.82	65.22**	.59		.14		
Reading accuracy ¹	14	90.71	14.95	14	90.29	16.02	14	89.40	13.56	19	90.05	14.99	19	92.32	12.77	17	91.06	12.24	.81	63.79**	.58		.22		
Reading comprehension ¹	14	88.93	14.40	14	87.57	13.94	14	88.27	12.00	19	89.05	15.55	19	88.16	13.85	17	89.18	12.75	.68	32.46**	.05		.00		
Reading rate ¹	14	106.57	13.70	14	109.43	16.70	14	102.53	13.75	19	105.11	15.61	19	108.26	10.07	17	109.06	12.07	.33	6.98**	1.76		.66		
Visual scanning and attention																									
Visual scanning																									
Items correct ⁴	15	1.84	.59	15	1.48	.40	15	1.29	.31	18	1.80	.71	19	1.59	.51	9	1.22	.43	.75	31.88***	1.43		.33		
Attention																									
Omissions ³	15	63.82	23.33	15	59.38	14.55	15	58.88	12.37	19	61.71	16.41	19	60.93	14.07	16	59.80	17.82	.22	3.90*	.13		.03		
Commissions ³	15	49.49	8.58	15	53.32	8.49	15	50.45	12.65	19	53.16	5.06	19	55.18	7.02	16	53.46	9.38	.20	3.44*	.21		.03		
Hit Rate ³	15	62.36	11.43	15	58.72	11.76	15	59.51	14.51	19	56.28	14.97	19	55.14	13.12	16	52.26	11.10	.19	3.24*	1.38		.28		

¹Standard score. ²Trials correct. ³*t*-scores. ⁴Seconds. * $p < .05$; ** $p < .01$; *** $p < .001$.

Dahlin, 2010). This is the first indication using an RCT methodology that this relatively short but intensive intervention can lead to long-term improvements in verbal WM, the system that is likely to support learning in the predominantly verbal environment of the classroom. The effect size for the gains in verbal WM was large, though due to the relatively low power it would be premature to draw strong conclusions about the specificity of the long-term gains.

In an earlier study by Holmes *et al.* (2009), WM training was associated with improvements in the abilities of children with low WM to follow multi-step spoken instructions. In contrast, there were no significant enhancements to performance on laboratory analogues of WM-related classroom tasks, including following instructions, immediately after training in the present study. The only change observed in these tasks was in processing ability at follow-up where the accuracy with which children counted the number of words in the sentence counting task improved; there was no significant enhancement in the children's recall performance on this task. This represents a major limitation on the utility of training as, clearly, its potential value is in enhancing classroom functioning rather than performance on laboratory-based WM assessments. One possibility is that the benefits of training are simply restricted to computer-based WM tasks that share many of the surface features of the training tasks (Dahlin, Neeley, Larsson, Backman & Nyberg, 2008). Alternatively, it may be that the training regime employed here only does half of the job required. One of the cardinal principles of neurorehabilitation is that scaffolding and support is required for training to generalize and be effective in new situations (Wilson, 2008). WM trainees may therefore need guidance, practice and reinforcement to apply their newly developed skills or strategies to everyday activities with structures that deviate substantially from the trained tasks but which nonetheless depend in part on WM.

The greatest improvements in WM following training were observed in complex span measures strongly associated with children's academic achievements in literacy and mathematics (Swanson & Siegel, 2001; Alloway, Gathercole, Willis & Adams, 2004). However, adaptive WM training did not significantly improve children's performance on standardized reading and mathematics tests either immediately after training or one year later. Indeed, the only significant change in any group was an increase in basic reading scores for the no intervention group. These data stand in contrast with Holmes *et al.*'s (2009) findings that mathematical abilities were enhanced by adaptive training 6 months after training was completed. However, this earlier study

lacked the comparison control condition at follow-up required to provide a stringent test of the specificity of training gains. It does, however, remain possible that outcome measures employed to date lack sufficient sensitivity to detect subtle and developing changes in learning abilities. In this respect, process-based tests of aspects of reading and mathematics abilities known to tax WM (e.g. Adams & Hitch, 1997; Geary, Hoard, Byrd-Craven, Nugent & Numtee, 2007) may well provide more sensitive measures of changes in WM capacity *in situ*.

Training had no significant impact on visual scanning or the ability to sustain attention over extended periods. It also had no effect on nonverbal reasoning, contrary to studies that have used *n*-back training paradigms (e.g. Jaeggi, Buschkuhl, Jonides & Perrig, 2008; Jaeggi *et al.*, 2011), and others in which CWMT (Klingberg *et al.*, 2005) has been used despite comparable statistical power. We therefore have no evidence to support claims that WM training enhances non-verbal IQ.

In summary, this study establishes two important facts about WM training using a stringent methodology designed to satisfy critics of the field. First, training in low WM children leads to generalized enhancements to a wide range of untrained WM tasks. Second, these gains do not translate into capacity improvements on ecologically valid measures of WM or to gains in academic progress. A priority now is to establish whether additional training activities can be developed to promote the flexible application of newly enhanced WM skills to less predictable memory-demanding situations in the classroom.

Acknowledgements

This research was supported by grant R1165301 to Susan Gathercole from the Leverhulme Trust. We would like to thank the children, teachers and parents whose participation made the study possible and Pearson for allowing us access to the Cogmed training program for the purposes of this evaluation.

References

- Adams, J.W., & Hitch, G.J. (1997). Working memory and children's mental addition. *Journal of Experimental Child Psychology*, *67*, 21–38.
- Alloway, T.P. (2007). *Automated Working Memory Assessment*. London: Pearson Assessment.
- Alloway, T.P., Gathercole, S.E., & Pickering, S.J. (2006). Verbal and visuo-spatial short-term and working memory in

- children: are they separable? *Child Development*, **77**, 1698–1716.
- Alloway, T.P., Gathercole, S.E., Willis, C., & Adams, A.-M. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology*, **87**, 85–106.
- Baddeley, A.D., Gathercole, S.E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, **105**, 158–173.
- Cain, K., Oakhill, J., & Bryant, P.E. (2004). Children's reading comprehension ability: concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, **96**, 31–42.
- Cogmed Working Memory Training (2011). [online]. Available at: <http://www.cogmed.com/rm> (Accessed 22 February 2013).
- Conners, C.K., & Multi-Health Systems Staff (2004). *Continuous Performance Test II*. Toronto: Multi-Health Systems Inc.
- Dahlin, E., Neeley, A.S., Larsson, A., Backman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science*, **320**, 1510–1512.
- Dahlin, K.I.E. (2010). Effects of working memory training on reading in children with special needs. *Reading and Writing*, **24**, 479–491.
- Delis, D.C., Kaplan, E., & Kramer, J.H. (2001). *Delis-Kaplan Executive Function System (D-KEFS)*. San Antonio, TX: Psychological Corporation.
- Engle, R.W., Carullo, J.J., & Collins, K.W. (1991). Individual differences in working memory for comprehension and following directions. *Journal of Educational Research*, **84**, 253–262.
- Erdfeiler, E., Faul, F., & Buchner, A. (1996). GPOWER: a general power analysis program. *Behavior Research Methods, Instruments, & Computers*, **28**, 1–11.
- Gathercole, S.E., & Alloway, T.P. (2008). *Working memory and learning: A teacher's guide*. London: Sage Publications.
- Gathercole, S.E., Alloway, T.P., Kirkwood, H.J., Elliott, J.E., Holmes, J., & Hilton, K. (2008a). Attentional and executive function behaviours in children with poor working memory. *Learning and Individual Differences*, **18**, 214–223.
- Gathercole, S.E., Brown, L., & Pickering, S.J. (2003). Working memory assessments at school entry as longitudinal predictors of National Curriculum attainment levels. *Education and Child Psychology*, **20**, 109–122.
- Gathercole, S.E., Dunning, D.L., & Holmes, J. (2012). Cogmed training: let's be realistic about intervention research. *Journal of Applied Research in Memory and Cognition*, **1**, 201–203.
- Gathercole, S.E., Durling, M., Evans, S., Jeffcock, E., & Stone, S. (2008b). Working memory abilities and children's performance in laboratory analogues of classroom activities. *Applied Cognitive Psychology*, **22**, 1019–1037.
- Gathercole, S.E., Lamont, E., & Alloway, T. (2006). Working memory in the classroom. In S.J. Pickering (Ed.), *Working memory and education* (pp. 219–249). Oxford: Elsevier Press.
- Geary, D.C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities*, **37**, 4–15.
- Geary, D.C., Hoard, M.K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development*, **78**, 1343–1359.
- 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., & Dunning, D.L. (2010a). Poor working memory: impact and Interventions. In P. Bauer (Ed.), *Advances in Child Development and Behaviour, Volume 39*; J. Holmes (Ed.), *Developmental Disorders and Interventions* (pp. 1–43). London: Academic Press.
- Holmes, J., Gathercole, S.E., Place, M., Dunning, D.L., Hilton, K.A., & Elliott, J.G. (2010b). Working memory deficits can be overcome: impacts of training and medication on working memory in children with ADHD. *Applied Cognitive Psychology*, **24**, 827–836.
- Jaeggi, S.M., Buschkuhl, 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**, 6829–6833.
- Jaeggi, S.M., Buschkuhl, 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.
- Kaufman, A.S., & Kaufman, N.L. (2004). *Kaufman Test of Educational Achievement*. Minneapolis, MN: Pearson Assessments.
- Kerry, S.M., & Bland, J.M. (1998). The intraclass correlation coefficient in cluster randomisation. *BMJ*, **316**, 1455.
- Klingberg, T. (2012). Is working memory capacity fixed? *Journal of Applied Research in Memory and Cognition*, **3**, 194–196.
- Klingberg, T., Fernell, E., Olesen, P.J., Johnson, M., Gustafsson, P., Dahlstrom, K., Gillberg, C.G., Forsberg, H., & Westerberg, H. (2005). Computerized training of working memory in children with ADHD – a randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, **44**, 177–186.
- Klingberg, T., Forsberg, H., & Westerberg H. (2002). Increased brain activity in frontal and parietal cortex underlies the development of visuospatial working memory capacity during childhood. *Journal of Cognitive Neuroscience*, **14**, 1–10.
- Koch, G.G. (1982). Intraclass correlation coefficient. In S. Kotz & N.L. Johnson (Eds.), *Encyclopedia of statistical sciences* (pp. 213–217). New York: John Wiley & Sons.
- Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159–174.
- McCarney, R., Warner, J., Iliffe, S., van Haselen, R., Griffin, M., & Fisher, P. (2007). The Hawthorne Effect: a randomised, controlled trial. *BMC Medical Research Methodology*, **7**, 30.
- Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, **49**, 270–291.
- Mezzacappa, E., & Buckner, J. (2010). Working memory training for children with attention problems or hyperactivity:

- a school-based pilot study. *School Mental Health*, **2** (4), 202–208.
- Neale, M.D. (1997). *Neale Analysis of Reading Ability (NARA)*. London: GL Assessment.
- Oberauer, K., Suß, K.-M., Wilhelm, O., & Wittman, W.W. (2008). Which working memory functions predict intelligence? *Intelligence*, **36**, 641–652.
- Shipstead, Z., Redick, T.S., & Engle, R.W. (2010). Does working memory training generalize? *Psychologica Belgica*, **50**, 245–276.
- Shipstead, Z., Redick, T.S., & Engle, R.W. (2012). Is working memory training effective? *Psychological Bulletin*, **138**, 628–654.
- Swanson, H.L., & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: both executive and phonological processes are important. *Journal of Experimental Child Psychology*, **79**, 294–321.
- Swanson, H.L., & Siegel, L. (2001). Learning disabilities as a working memory deficit. *Issues in Education: Contributions of Educational Psychology*, **7** (1), 1–48.
- Torgerson, D.J., & Torgerson, C.J. (2008). *Designing randomised trials*. Chippenham and Eastbourne: Palgrave Macmillan.
- Wechsler D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. London: Harcourt Assessment.
- Wechsler, D. (1996). *Wechsler Objective Number Dimensions (WOND)*. New York: Psychological Corporation.
- Wechsler, D. (1993). *Wechsler Objective Reading Dimensions (WORD)*. New York: Psychological Corporation.
- Wilson, B.A. (2008). Neuropsychological rehabilitation. *Annual Review of Clinical Psychology*, **4**, 141–162.

Received: 9 March 2012

Accepted: 7 March 2013