Testing for near and far transfer effects with a short, face-to-face adaptive working memory training intervention in typical children

Key words: working memory, training, intervention, children, reading comprehension

Running Head: working memory training intervention

Abstract (200 words)

A relatively quick, face-to-face, adaptive working memory training intervention was assessed in 5- to 8-year-old typically developing children, randomly allocated to a six-week intervention condition, or an active control condition. All children received 18 sessions of 10 minutes, three times/week for six weeks. Assessments of six working memory skills, word reading and mathematics were administered at pre-test, post-test, and six month follow-up. Additional measures of word reading, mathematics, spelling and reading comprehension were given at a 12 month follow-up. At post-test, the trained group showed significantly larger gains than the control group on the two trained executive-loaded working memory tasks (Listening Recall, Odd One Out Span) and on two untrained working memory tasks (Word Recall, Counting Recall). These "near transfer" effects were still apparent at six month follow-up. "Far transfer" effects were less evident: there was no difference between the groups in their gains on single word reading and mathematics over 12 months, and spelling skills did not differ at 12 month follow-up. However, the trained group showed significantly higher reading comprehension scores than the control group at 12 month follow-up. Thus, improving the ability to divide attention between processing and storage may have had specific benefits for reading comprehension.

The purpose of the current study was to assess the effectiveness of a relatively short, adaptive, face-to-face working memory (WM) training intervention delivered three times a week, for about 10 minutes, over a period of six weeks to typically-developing children between the ages of five and eight years. In order to ensure methodological adequacy, participants were randomly assigned to a treated control group or to a training intervention group; and near and far transfer effects were investigated by assessing working memory and academic skills in children before training, after training, and at six months follow-up. Selected measures were also administered at a final 12 month follow-up. The context for this work was the ongoing debate about whether working memory training interventions might prove valuable for children with low working memory and various developmental disorders. Below, we highlight the key methodological and theoretical issues that informed the work.

In a recent meta-analytic review of WM training interventions, Melby-Lervåg and Hulme (2012) concluded that such interventions produce reliable short-term improvements ("near-transfer" effects), particularly for children under ten years, but that these are not durable (after around nine months no effects were detectable). Further, they found no evidence that WM training produced improvements in other cognitive skills, including non-verbal and verbal ability, arithmetic, word decoding, and inhibitory processes in attention ("far-transfer" effects). Melby-Lervåg and Hulme (2012) argued, therefore, that such interventions cannot currently be recommended as treatment options for children with low working memory or developmental disorders associated with working memory problems (e.g., language difficulties, ADHD, mathematical difficulties, dyslexia - for relevant research see Archibald & Gathercole, 2006; Ellis Weismer, Evans & Hesketh, 1999; Martinussen, Hayden, Hogg-Johnson &Tannock, 2005; Passolunghi, 2006; Swanson, 2006), particularly when other interventions that directly target areas of weakness may have more reliable effects (e.g., direct training of reading and language skills in children with Down syndrome, Burgoyne et al., 2012; language interventions for children with developmental language difficulties, Ebbels, 2007). There also are methodological issues with some of the training studies, Melby-Lervåg and Hulme (2012) and Shipstead et al. (2012) noted that not all studies have included random assignment of participants to trained and untrained groups, adequate pre-and post-testing, active control groups, or a sufficient range and depth of outcome measures. Shipstead, Redick and Engle (2012), in another recent review, cautioned that there is no compelling evidence to link WM training *directly* to gains in working memory capacity, and that evidence of far transfer effects must be clearly demonstrated.

Both reviews also identified a number of theoretical issues that require attention. For example, the concept of 'executive-loaded' working memory (ELWM) needs to be clearly specified, as this reflects the ability to process and store information concurrently (Baddeley, 2000; Case, Kurland & Goldberg, 1982; Daneman & Carpenter, 1980), as distinct from short-term memory (STM), which is the ability temporarily to store and recall verbatim sets of verbal (e.g., digits, words) or visual/spatial items (e.g., visual details, spatial positions) (Baddeley, 2000). There is little positive evidence for durable training and transfer effects from interventions involving verbal STM (for a review see Kail, 1990). Consequently, WM training programs have focused on training ELWM as this is regarded as a critical cognitive skill (e.g. Klingberg, 2010). For example, Melby-Lervåg and Hulme (2012) noted that the dominant view in the literature is that ELWM represents "a domain general attentional resource limitation" (page 2) closely linked to fluid intelligence (e.g., Friedman et al., 2006; see also Shipstead, Redick & Engle, 2012). However, it is not clear exactly how or why training this skill should improve capacity; current explanations do not move beyond concepts around strengthening or stretching a child's capacity through repeated practice.

The purpose of the current study was to address as many of these issues as possible, and to introduce a non-computer-based face-to-face intervention that was quick and easy to administer. Participants were typical children with average abilities and they were randomly allocated to a training intervention or a treated control group. The training intervention involved repeated practice with two ELWM tasks, both requiring concurrent processing and storage, in an *adaptive* manner such that span administration levels were related to current performance. Sessions were considerably shorter than the 30-60 minutes suggested by Klingberg, (2010), but the total number of sessions (18) conformed closely to his recommendations (20 sessions). The two EWLM tasks were listening span (processing = judge the veracity of sentence/s, storage = recall the final word/s of the sentence/s), and odd one out span (processing = point to the odd one/s out of three nonsense shapes, storage = recall their spatial position/s). Participants in the treated control group carried out just the processing parts of each task, i.e., they made the judgments about sentences or odd ones out, but there was no memory/storage requirement. Hence, both groups had equal exposure to the materials used in training and to an enthusiastic 'trainer', but only one group carried out the critical ELWM processing plus storage task. A range of working memory skills were assessed at three time points (pre, post, six month follow-up) using reliable (mainly standardised) tests to assess near transfer. Standardised tests of scholastic abilities (reading, mathematics) were administered at four time points (pre, post, six month and 12 month follow-up) to assess far transfer more thoroughly than in many previous studies.

We assessed several related issues. Firstly, would the intervention have: (1) direct benefits on the ELWM tasks we trained (a necessary pre-condition for an effective intervention); (2) near-transfer effects to working memory tasks we did not directly train; and (3) far-transfer effects to important scholastic abilities, i.e., reading and mathematics (following the designs of similar previous studies e.g., Holmes, Gathercole & Dunning, 2009; Loosli, Buschkuehl, Perrig & Jaeggi, 2012; St Clair-Thompson, Stevens, Hunt & Bolder, 2010; Van der Molen, Van Luit, Van der Molen, Klugkist & Jongmans, 2010). Secondly, would these potential training, near- and far-transfer effects be maintained after a period of six months, demonstrating the durability of the intervention.

In addition, we re-tested single word reading and mathematics at a 12 month follow-up assessment. As time resources were limited at this stage due to school availability, the ideal design (repeating ALL tests as well as including some new tests to broaden the scope of the study) was not possible. A decision was made to prioritise scholastic tests rather than repeat the entire battery of working memory measures. We, therefore, introduced two new tasks at this point, namely reading comprehension and single word spelling, neither of which had been assessed in previous studies. It is important to note that this represents both a strength and a limitation of the study design. Current debates over the durability of WM interventions (Melby-Lervåg & Hulme, 2012) mean that it would have been useful to obtain information about working memory at a 12 month follow-up. Yet, we felt it was more important for the literature on WM interventions to look further afield for potential far transfer effects as this would provide the most convincing evidence of the value of such interventions. Furthermore, broadening the search for far transfer effects had the potential to shed more light on mechanisms of change underlying any improvements in performance from WM training interventions.

The hypothesis in relation to far transfer effects on reading comprehension was that this skill might improve *some time after* successful WM training via increasing a child’s capacity for continuous monitoring and updating of the contents of working memory, which is a key skill required for successful comprehension (e.g., Hulme & Snowling, 2009). Related to this are well-established relationships between reading comprehension and ELWM abilities in children (e.g., Cain, Oakhill & Bryant, 2004; Leather & Henry, 1994; see also Nation, Adams, Bowyer-Crane & Snowling, 1999, for similar findings on children with reading comprehension difficulties) and adults (Daneman & Carpenter, 1980; Engle, Tuholski, Laughlin & Conway, 1999). If this hypothesis were to be confirmed, the findings would suggest a possible mechanism underlying WM training benefits, i.e., via improvements in the ability to choose and retain relevant information whilst suppressing other information not currently required. Predictions in relation to spelling were less well-specified, but we hypothesised that the serial generation and output demands of spelling might make greater demands on ELWM than single word reading, as it is a more developmentally advanced task involving retaining the target word while simultaneously producing individual letters and checking of spelling patterns. Thus, it was hoped that rather than simply 'exercising' ELWM (Melby-Lervåg & Hulme, 2012), this training intervention would target and improve a key cognitive skill that is relevant to scholastic and everyday abilities, so that over a period of a year children would show gains in 'far' abilities.

Method

Participants and Selection

Recruitment letters were sent out to all parents/carers of children in year groups one and three, in one Greater London school, with a view to recruiting 18 children from each year group. Letters explained that once consent was given, the children would be randomly assigned to either the 'training' or the 'control' group. Out of 115 letters sent, there were 52 positive responses, within the one week deadline, meaning that the study was oversubscribed. The school selected out those they regarded as the very brightest/weakest children, leaving 36 children (mean age in months = 84m, SD 12.94m, range 67m - 102m) to take part in the study. All participants had verbal and non-verbal abilities within the normal range, with a good range of ability within the cohort. Children were randomly assigned to the Training group (n=18, mean age 84.2m, SD 13.1m, Mean BAS-II Matrices T-Score 57.6, SD 6.2, Mean BAS-II Verbal Similarities T-Score 53.8, SD 6.0) or the Control group (n=18, mean age 84.0m, SD = 13.1m, Mean BAS-II Matrices T-Score 59.7, SD 7.5, Mean BAS-II Verbal Similarities T-Score 54.4, SD 4.7). The only limiting factors affecting randomisation were year group (year one, year three) and gender to ensure equal distribution across groups.

There were no significant differences between the groups at pre-test for age ($t(24) = 0.04$, n.s.), BAS-II Matrices T-Scores ($t(34) = -0.95$, n.s.), or BAS-II Verbal Similarities T-Scores ($t(34) = -.34$, n.s.). (No differences between groups were present for any other variable either, as will be described in the results section below). At post-test / six month follow-up, one participant in the Control group had moved away (n=17); at 12 month follow-up a second participant in the Control group had moved away (n=16).

This project was granted ethical approval from the Research Ethics Committee, London South Bank University, and was discussed in detail with appropriate school staff before recruitment. Informed consent for participation was obtained in writing from parents/guardians; children/students also gave their written consent and were told they could opt out at any time. Testing and training took place at school, in the same place for every session, and always with the same Experimenter (GN).

<u>Assessments</u>

Participants were assessed four times: at pre-test before our intervention; at post-test after our intervention; at six month follow-up; and at 12-month follow-up. Figure 1 shows the tests that were administered at each time point.

Figure 1 about here

General Verbal and Nonverbal Ability (IQ)

Two measures from The British Abilities Scales (BAS-II, Elliott, Smith & McCullouch, 1996) were used. *Verbal Similarities* - the child was presented with three words and asked to state how those three things were similar or went together (e.g., banana, apple, orange – they are all fruit). *Matrices* - the child was shown an incomplete matrix of abstract figures, and they had to select, from six choices, the figure that correctly completed the matrix.

Working Memory Tests

All of the working memory tests, except one (Odd One Out Span), were standardised measures from the Working Memory Test Battery for Children (WMTB-C, Pickering & Gathercole, 2001). Four tests remained 'untrained' for the duration of the study (Digit Recall, Word Recall, Block Recall, Counting Recall), and two tests were trained using similar (but not identical) materials (Listening Recall, Odd One Out Span).

Outcome Measures Working Memory ('Near Transfer')

Digit Recall - the experimenter read a list of numbers and the child had to repeat the numbers back immediately, in the correct order. Trials began with lists of single numbers and were given in blocks of six trials per list length; the list length was increased until the child made three errors or more within a block. Total trials correct were scored.

W*ord Recall* - the experimenter read a list of single syllable words and the child had to repeat the words back immediately, in the correct order. Trials began with lists of single words and were given in blocks of six trials per list length; the list length was increased until the child made three errors or more within a block. Total trials correct were scored.

Block Recall - the experimenter placed the Block board between themselves and the child at a comfortable distance from both, and so that the numbers face the examiner only and cannot be seen by the child. The experimenter tapped out sequences of spatial positions on the blocks and the child was asked to immediately touch exactly the same blocks in the same sequence. Trials began with lists of single items and were given in blocks of six trials per list length; the list length was increased until the child made three errors or more within a block. Total trials correct were scored.

Counting Recall - the experimenter asked the child to count, out loud, the number of dots presented on a page of the stimuli book. The page was turned and then the child was asked to recall the number of dots on the page. Again, trials started with single pages of dots, and were given in blocks of six trials per list length; the list length was increased, until the child made three or more errors within a block. The child had to count items out loud for each page, but not recall the 'outcome' numbers until after all the pages in that trial had been counted and turned over. Numbers had to be recalled in the order in which they were counted. Total trials correct were recorded.

Two further tests were both *assessed* and also *adapted* for the WM training intervention. Note that performance on these tests was assessed using *independent* measures at each time point. Although the WM interventions used the same format as the independent tests, they contained entirely different stimuli.

Listening Recall - the experimenter read sequences of short sentences, which did, or did not make sense (e.g., ‘lions have four legs’, or ‘pineapples play football’). The child was asked to say whether the sentence was ‘true’ or ‘false’. Then the child was asked to recall the final word of the sentence. As above, trials started with single item sequences and increased, in blocks of six, until there were three errors or more made within a block. The child was asked to judge the veracity of each sentence in a trial, immediately after it was spoken, but not to recall the final word of each sentence until after the total number of sentences in that trial (ranging from 1 to 6) had been heard. It was important that these target words were recalled in the order in which they were presented by the examiner. Total trials correct were scored but threshold span levels were also noted for the WM training intervention ‘start point’.

Odd One Out Span (Henry, 2001) - This task was designed to assess a child’s ELWM in the nonverbal, visuo-spatial domain. The test consisted of 63 cards (20x4cm), each showing three shapes. On each card, two of these shapes were identical and one was slightly different (the odd one out). The examiner showed the child one card and asked them to point to the odd one out. The card was then turned over, and a response sheet displaying a blank card of similar shape and size, with three squares, but with NO shapes, was presented. The child was asked to point, on this sheet, to the location (left, middle or right) of the odd one out on the card they had just seen. If the child verbalised an answer, they were asked only to point and to say nothing throughout the test. After three single-item trials, the child was shown two cards, one after the other, and was asked to point to the odd one out, on each. Then a sheet showing two blank cards was presented and the child was asked to point to the locations of the odd ones out, on each of the shape cards they had just seen, in the same order as they were presented. Trial list lengths increased in blocks of three until the child made two or more errors within a block. Total trials correct were recorded, but threshold span levels were also noted for the WM training intervention 'start point'.

Outcome Measures ('Far Transfer')

Two measures from BAS-II (Elliott et al., 1996) were used. *Number Skills* - the child performed various number-based tasks, such as pointing to orally administered numbers, naming visually presented numbers and doing written calculations. *Word Reading* - the child read aloud a series of words presented on a card. These two measures were administered at every test point during the study.

Two further measures were included only at the 12-month follow-up. From the BAS-II (Elliott et al., 1996), *Spelling* – the child was asked to spell increasingly difficult single words with a sentence context. From the Wechsler Objective Reading Dimensions (WORD, Wechsler, 1993), we assessed *Reading Comprehension*. Here, the child was asked to read short, but increasingly difficult passages of text, and then answer a question about what they had read. This task has been regarded as assessing the child's ability to make elaborative inferences about a text (Hulme & Snowling, 2009).

Working Memory Training Intervention (6 weeks)

Each child met with the experimenter, three times a week, for about 10 minutes. The Training Group were given the two ELWM intervention tasks, the Control Group received equal one to one attention, but did simpler versions of the tasks, requiring only the processing part of each task, with no requirement for memory storage. Exactly the same materials were used with both groups, and these were entirely new items that were developed based on the Listening Recall and the Odd One Out tasks.

The intervention comprised adaptive practice on both the Listening Recall and the Odd One Out tasks. Each child's span level for both tasks had been established during the pre-training assessments. This was used to make the training adaptive and therefore individually matched to each child's ability. No child was made to feel that the training was beyond their capabilities for long periods of time because correct performance led to more difficult (longer) lists being presented, whereas incorrect performance led to easier (shorter) lists being presented.

A training session consisted of 11 trials of the listening recall and 11 trials of the odd-one-out test. Trials were administered, starting at the span established for each child at pre-test. Children with spans of just one item started their training trials on two items automatically, as trials of one item were not regarded as executive-loaded. If two consecutive trials were responded to correctly, the span level was increased by one for the next two trials; similarly, if two consecutive trials were responded to incorrectly the span level for the next two trials was decreased by one. This procedure was employed for all 11 trials on both tasks during each intervention session, after which the experimenter judged the spans for that entire session, based on the most frequently achieved correct span length for each task. These span lengths were used at the start point for the next training session, which took the same format as described above, and so on until each child had received 18 training sessions over a period of 6 weeks (note that in some cases due to absence the 18 training sessions were delivered over 7 or 8 weeks as required). All participating children received praise and brightly coloured stickers for taking part.

Active Control Condition (6 weeks)

These sessions were the same length and frequency as the intervention sessions, but children simply made judgements about sentence veracity or odd one out locations (i.e., 'processing'), and there was no memory requirement.

For both types of training the children enjoyed the sessions, almost without exception. They were short enough for children not to get bored or to feel threatened, and difficult enough to present an achievable challenge, which they delighted in. The one-to-one aspect of the sessions appeared vital in boosting children's confidence and willingness to continue sessions. They often had stories to tell and news to share about other things going on in their lives and appreciated a chance to chat. Many of the children particularly liked the 'silly sentences' as they called them, and often tried to make some up. They sometimes talked about practising the training at home and trying both the tasks out on their parents. Some children actually came along with some sentences that they had written themselves and the teachers were thrilled with this initiative.

Results

Table 1 shows mean scores on all study measures at all relevant time points, given in terms of both raw/ability scores and standard scores. We also include graphical representations of training gains for raw scores (see Figure 2). Analyses of variance were always carried out on raw scores for ease of comparison between testing points (standard scores sometimes give a skewed indication of performance for individual children who have just crossed a new age band or have just remained in a previous age band). These analyses were carried out for our two ‘trained’ ELWM measures, as well as for each of the ‘near transfer’ (other working memory skills) and ‘far transfer’ (scholastic abilities) outcome measures. The analyses comprised one between subjects factor (training vs. control) and one within subjects factor (time point – note that the number of *levels* for time point varied between tasks, see Figure 1). Where violations of sphericity occurred, Greenhouse-Geisser corrections were applied. Follow-up planned *t*-tests were carried out to examine group differences at each time point when significant interactions were found between Time and Group, with Cohen’s *d*-values reported where relevant.

Table 1 about here

Performance on Trained Measures (Listening Recall, Odd One Out Span: Pre-test, Post-test, 6m Follow-up)

It is a necessary pre-condition of a training intervention that the trained skill improves, and this was confirmed for both trained measures. For Listening Recall, the ANOVA indicated a significant effect of Time, $F(2, 66) = 32.71$, $p < .001$, partial eta^2 = .498; a significant effect for Group, $F(1, 33) = 35.51$, $p < .001$, partial eta^2 = .518; and, critically, an interaction between Time x Group, $F(2, 66) = 13.55$, $p < .001$, partial eta^2 = .291. Planned comparisons indicated that the two groups did not differ at pre-test ($t(34) = 1.36$, n.s.), but that they differed significantly at post-test ($t(34) = 5.36$, $p < .001$, $d = 1.81$) and at follow-up ($t(33) = 6.35$, $p < .001$, $d = 2.18$), with the trained group having better Listening Recall performance in both cases. Therefore, listening recall performance improved more in the trained than the control group.

Subsequent one-way ANOVAS were used to explore the interaction by examining changes in Listening Recall performance at each assessment point for each group separately. There was a significant effect of Time in the trained group, $F(1.49, 25.40) = 33.95$, $p < .001$, partial eta^2 = . 666. Paired contrasts indicated that post-test performance exceeded pre-test performance, $F(1, 17) = 36.37$, $p < .001$, partial eta^2 = . 681; and that six month follow-up performance exceeded pre-test performance, $F(1, 17) = 53.56$, $p < .001$, partial eta^2 = .759. There was also a significant effect of Time in the control group, $F(2, 32) = 3.65$, $p < .001$, partial eta^2 = .186. Paired contrasts indicated that post-test performance exceeded pre-test performance, $F(1, 16) = 5.36$, $p < .05$, partial eta^2 = .251; and that six month follow-up performance exceeded pre-test performance, $F(1, 16) = 5.87$, $p < .05$, partial eta^2 = .268. Therefore, significant improvements in performance compared to pre-test were observed in both groups. However, looking at standard scores (Table 1) gives an indication of the relative size of the increases in Listening Recall in both groups. In the trained group, scores improved from 100 at pre-test to 130 at post-test, then dropped slightly to 123 at the six month follow-up. Increases in the control group were smaller: from 94 (pre-test) to 99 at post-test and 96 at six month follow-up.

For Odd One Out Span, the ANOVA indicated a significant effect of Time, $F(1.56, 51.52) = 42.59$, $p < .001$, partial eta^2 = .563; a significant effect for Group, $F(1, 33) = 14.66$, $p < .01$, partial eta^2 = .308; and, critically, an interaction between Time x Group, $F(1.56, 51.52) = 12.49$, $p < .001$, partial eta^2 = .275. Planned comparisons indicated that the two groups did not differ at pre-test ($t(34) = 0.61$, n.s.), but that they differed significantly at post-test ($t(34) = 3.82$, $p < .01$, $d = 1.30$) and at six month follow-up ($t(33) = 4.37$, $p < .001$, $d = 1.51$), with better Odd One Out Span performance in the trained group in both cases. Therefore, Odd One Out Span performance improved more in the trained than the control group.

Subsequent one-way ANOVAS were used to explore the interaction by examining changes in performance at each assessment point for each group separately. There was a significant effect of Time in the trained group, $F(2, 34) = 44.52$, $p < .001$, partial eta^2 = .724. Paired contrasts indicated that post-test performance exceeded pre-test performance, $F(1, 17) = 53.95$, $p < .001$, partial eta^2 = .760; and that six month follow-up performance exceeded pre-test performance, $F(1, 17) = 54.09$, $p < .001$, partial eta^2 = .761. There was also a significant effect of Time in the control group, $F(1.48, 23.66) = 5.46$, $p < .05$, partial eta^2 = .255. Paired contrasts indicated that post-test performance exceeded pre-test performance, $F(1, 16) = 5.67$, $p < .05$, partial eta^2 = .262; and that six month follow-up performance exceeded pre-test performance, $F(1, 16) = 6.67$, $p < .05$, partial eta^2 = .294. Significant improvements in performance compared to pre-test were observed in both groups, but were larger in the trained group.

Near transfer effects (Digit, Word, Block and Counting Recall: Pre-test, Post-test, 6m Follow-up)

The ANOVA for Digit Recall indicated a significant effect of Time, $F(1.67, 55.02:) = 31.05$, $p < .001$ partial eta^2 = .485, but no significant effect for Group and no interaction between Time and Group. Digit Recall improved over the six month period for both groups to the same extent.

The ANOVA for Word Recall indicated a significant effect of Time, $F(2, 66) = 28.38$, $p < .001$, partial eta^2 = .462; a significant effect for Group, $F(1, 33) = 20.84$, $p < .001$, partial eta^2 = .387; and, critically, an interaction between Time x Group, $F(2, 66) = 6.76$, $p < .01$, partial eta^2 = .170. Planned comparisons indicated that the two groups did not differ at pre-test

($t(34) = 1.51$, n.s.), but that they differed significantly at post-test ($t(34) = 3.29$, $p < .01$, $d = 1.11$) and at six month follow-up ($t(33) = 5.77$, $p < .001$, $d = 1.98$), with better Word Recall performance in the trained group. Therefore, Word Recall performance improved more in the trained than the control group.

Subsequent one-way ANOVAS to examine the changes in performance at each assessment point for each group separately revealed a significant effect of Time in the trained group, $F(2, 34) = 28.50$, $p < .001$, partial eta^2 = .626. Paired contrasts indicated that post-test performance exceeded pre-test performance, $F(1, 17) = 34.46$, $p < .001$, partial eta^2 = .670; and that six month follow-up performance exceeded pre-test performance, $F(1, 17) = 56.48$, $p < .001$, partial eta^2 = .769. There was also a significant effect of Time in the control group, $F(2, 32) = 4.29$, $p < .05$, partial eta^2 = .212. Paired contrasts indicated that post-test performance exceeded pre-test performance, $F(1, 16) = 5.64$, $p < .05$, partial eta^2 = .261; and that six month follow-up performance exceeded pre-test performance, $F(1, 16) = 11.73$, $p < .01$, partial eta^2 = .423. Significant improvements in performance compared to pre-test were observed in both groups. Looking at the size of the increases in Word Recall in terms of standard scores in the trained group (see Table 1) revealed that scores improved from 104 at pre-test to 120 at post-test, and that performance gains were maintained (score = 121) at the six month follow-up. Increases in the control group were smaller: from 99 (pre-test) to 105 (post-test), and down slightly to 101 at the six month follow-up.

The ANOVA for Block Recall indicated a significant effect of Time, $F(2, 66) = 4.68$, $p < .05$, partial eta^2 = .119; but no significant effect for Group and no interaction between Time and Group. Block Recall improved over the six month period for both groups to the same extent.

The ANOVA for Counting Recall indicated a significant effect of Time, $F(1.71, 56.36:) = 14.99$, $p < .001$, partial eta^2 = .312; a marginally significant effect for Group, $F(1, 33) = 3.86$, $p = .058$, partial eta^2 = .105; and an interaction between Time x Group, $F(1.71, 56.36) = 5.32$, $p < .05$, partial eta^2 = .139. Planned comparisons indicated that the two groups did not differ at pre-test ($t(34) = -.51$, n.s.), but that they differed significantly at post-test ($t(34) = 2.58$, $p < .05$, $d = 0.87$) and at follow-up ($t(33) = 2.29$, $p < .05$, $d = 0.79$), with better Counting Recall performance in the trained group. Therefore, Counting Recall performance improved more in the trained than the control group.

Subsequent one-way ANOVAS examined the changes in performance at each assessment point for each group separately. There was a significant effect of Time in the trained group, $F(1.50, 25.43) = 18.81$, $p < .001$, partial eta^2 = .525. Paired contrasts indicated that post-test performance exceeded pre-test performance, $F(1, 17) = 21.38$, $p < .001$, partial eta^2 = .557; and that six month follow-up performance exceeded pre-test performance, $F(1, 17) = 22.31$, $p < .001$, partial eta^2 = .568. There was no significant effect of Time in the control group, $F(2, 32) = 2.32$, n.s. Therefore, significant improvements in performance compared to pre-test were observed only in the trained group. Table 1 includes information on the size of the increases in Counting Recall in terms of standard scores: in the trained group scores improved from 89 at pre-test to 102 at post-test, and remained at 100 at the six month follow-up.

Far transfer effects (Word reading, Number Skills: Pre-test, Post-test, 6m Follow-up, 12m Follow-up) (Spelling, Reading Comprehension: 12m Follow-up)

The ANOVA for Word Reading ability scores indicated a significant effect of Time, $F(1.95, 62.35) = 69.47$, $p < .05$, partial eta^2 = .685; but no significant effect for Group and no interaction between Time and Group. Word Reading ability scores improved over the 12 month period for both groups to the same extent.

The ANOVA for Number Skills ability scores indicated a significant effect of Time, $F(3, 96) = 31.40$, $p < .05$, partial eta^2 = .495; but no significant effect for Group and no interaction between Time and Group. Number Skills ability scores improved over the 12 month period for both groups to the same extent.

Group differences in Reading Comprehension and Spelling were assessed using t-tests as we had data only from one time point (12-month follow-up). For Reading Comprehension, there were significant differences between the groups, favouring the trained group, $t(32) = 2.85$, $p < .01$, $d = 0.98$). Standard scores were higher in the trained group (109) than the control group (98). For Spelling, there were no significant differences between the groups ($t = 1.41$, n.s.).

Figure 2 about here

Discussion

The ‘short-session’, face-to-face adaptive WM training intervention assessed here was effective and durable. Children who were randomly allocated to a *processing plus storage* WM training intervention demonstrated significantly larger gains between pre-test and post-test in Listening Recall and Odd One Out Span performance, compared to a treated control group who had equal experience with *processing* the training materials, yet without the *storage* requirement. Expressing these improvements in terms of standard scores puts the findings into context: standard scores on Listening Recall improved from 100 to nearly 130 in the trained group, whereas improvements in the control group were small (from 94 to 99). Further, this relative advantage was maintained at a six month follow-up, demonstrating that the training effects were durable. This was necessary evidence for the efficacy of the WM training intervention, but it was not definitive because gains in performance could have been inflated by task-specific practice. More convincing evidence was provided by the near transfer effects from two untrained WM tasks: scores on Word Recall and Counting Recall improved significantly more in the trained group than in the control group between pre-and post test (improvements of 16 and 13 standard score points respectively), and these gains were also maintained at six month follow-up. Two measures of WM did not demonstrate the critical interaction between time of assessment and group that would have indicated near transfer effects, and for these measures (Digit Recall, Block Recall), both groups improved only slightly in performance over the period of the study.

These findings support previous reports in relation to near transfer effects, namely, that WM training interventions have beneficial effects on at least some other untrained WM skills in children (e.g., Alloway, Bibile & Lau, 2013; Gray et al., 2012; Holmes et al., 2009; Holmes et al., 2010; Klingberg et al., 2002; 2005; Loosli et al., 2012; Nutley et al., 2011; Thorell et al., 2009; Van der Molen et al., 2010). The six month follow up period was also longer than some earlier studies (Gray et al., 2012; Loosli et al., 2012; Van der Molen et al., 2010), and confirmed previous findings regarding the durability of WM improvements (Alloway et al., 2013; Holmes et al., 2009; Holmes et al., 2010). Of particular interest was the finding that adaptive training on two ELWM tasks (Listening Recall, Odd One Out) improved performance on a further measure of ELWM (Counting Recall) that employed a different processing task (counting dots) and a different storage task (recalling numbers). Near transfer effects on ELWM are argued to be of greater relevance than near transfer effects on STM, as the mechanisms for change via WM training interventions should reflect the specifically 'ELWM' requirement for divided attention between processing and storage (Shipstead et al., 2012).

It is worth considering why the short, face-to-face training intervention used here was so effective in improving WM scores. We focussed the training on executive-loaded WM because such tasks relate more closely to broader cognitive skills and achievements than simple STM storage tasks (Loosli et al., 2012). This may have made the training more generalisable to other WM tasks with executive loads (e.g., Counting Span) and aided performance on some similarly structured span tasks (e.g., Word Span). Further, the training sessions were short (about 10 minutes with a total of 22 trials of practice per session) and were given three days per week rather than daily, which is often done. The brevity of the sessions may have contributed to the excellent compliance we observed with our intervention. The most commonly used training intervention in the literature, COGMED, includes up to 10 different WM *and* STM skills and is lengthy if the entire package is used (e.g., 45 minutes training, 4-5 days a week for 5 weeks - Gray et al., 2012; or 35 minutes with 115 trials of practice in each session for 5-7 weeks - Holmes et al., 2009). Compliance levels are not usually described directly in the literature, although Gray et al. (2012) did note that a handful of participants dropped out of their intervention study using COGMED. Here, children enjoyed interacting with the 'live' experimenter, who could motivate and encourage them continuously to maintain interest and effort levels. Whilst other short WM training interventions (about 15 minutes per day) have generated positive effects on WM skills (Loosli et al., 2012; Nutley et al., 2011; Thorell et al., 2009), all used computer-based rather than face-to-face interventions. Both methods may be effective and, certainly, computer-based interventions are cost-effective. Finally, it should be noted that the active control condition in the current study differed from the non-adaptive WM training often employed in the literature. By equating time with the experimenter and experience with the WM training materials, the ‘processing only’ control condition ruled out non-specific effects of the intervention without requiring WM skills. This is important, as non-adaptive training improves WM skills to a minor extent (Holmes et al., 2009; Van der Molen et al., 2010), making interpretation of near and far transfer effects more difficult.

Despite the positive findings in respect of WM improvements, Shipstead et al. (2012) argue that near transfer effects following WM training interventions may only indicate that a child has adopted strategies to improve performance generally, or has benefitted from additional experience with certain types of stimuli. Hence, the improvements in Word Recall found here could be explained by practice in recalling lists of words in serial order during the Listening Recall portion of WM training. One potential mechanism of change could be the use of an explicit memory strategy such as chunking or verbal rehearsal that children discover to be useful across span tasks. Similarly, the improvements in Counting Recall could be explained by practice with similarly structured ELWM tasks that require concurrent serial recall of information whilst carrying out distracting processing tasks. It may be possible to carry out some form of rehearsal or chunking during ELWM tasks, or at least ‘reactivate’ the memory traces via rapid attentional switches (Barrouillet, Gavens, Vergauwe, Gaillard & Camos, 2009). As such, near transfer effects are promising and potentially important for the development and generalisation of new and effective memory strategies during childhood. Future work could include a larger range of WM measures to test these near transfer effects more thoroughly.

However, near transfer effects do not provide the strongest evidence for the efficacy of WM training interventions. Of more relevance are far transfer effects, as these demonstrate gains in abilities that are related to WM without being identical to WM tasks in either task structure or content. Reliable far transfer effects should indicate that *capacity* (or perhaps efficiency) has increased – otherwise there is no mechanism to account for the increases in performance on related tasks (Shipstead et al., 2012). Although we included several ability measures related to scholastic achievement to assess far transfer effects at several time points over the period of one year (single word reading, mathematics, spelling, reading comprehension), none demonstrated significantly greater gains in the trained than the treated control group. The exception was the measure of reading comprehension, which was assessed at one time point 12 months after the WM training intervention. This provided *suggestive* rather than *definitive* evidence that the intervention led to improvements in reading comprehension ability, but concurred with findings that 'executive-loaded' WM training improves reading comprehension in undergraduates (Chein & Morrison, 2010). For definitive evidence that WM improvements transfer to reading comprehension, the current design should have included the comprehension measure at all four time points during the study, particularly before the training intervention, to establish that there were no pre-existing group differences in comprehension. Nevertheless, it is important to note that the groups did not differ on any of the initial measures assessed before the WM training intervention and there was random allocation of children to the two conditions. As many of these measures (working memory, word reading, verbal ability, nonverbal ability) are themselves key predictors of reading comprehension (e.g., Cain et al., 2004: Leather & Henry, 1994), this provides at least some evidence that there were unlikely to have been significant 'pre-training' comprehension differences between the groups.

In terms of prior research findings with children, there is variability concerning far transfer effects to academic skills, with some finding beneficial effects on mathematics but not reading after 2 to 6 months (Holmes et al., 2009; Van der Molen et al., 2010), some finding improvements in spelling but not mathematics (Alloway et al., 2013), and others finding improvements in word and text reading immediately post-training (Loosli et al., 2012). However, few studies have looked for far transfer effects over a period of one year as was done here. Future work should consider including multiple measures of text reading, reading comprehension and language comprehension, administered at several time points over a period of at least a year, to test the generality the WM training intervention on both reading and comprehension skills that have heavy WM requirements (Loosli et al., 2012; Shipstead et al., 2012).

One further limitation of this study was that school availability did not permit us to test the full range of working memory measures again at 12-month follow-up. This would have been highly relevant to debates concerning the longevity of WM training interventions over periods greater than six months (Melby-Lervåg & Hulme, 2012). Future work should, ideally, include longer timescales in which to monitor gains in working memory performance and other academic skills. In addition, the ‘gold-standard’ for intervention studies incorporates double-blinding so that the experimenter conducting the WM intervention has no knowledge of group membership. It was not possible to achieve this in the current study, but we did ensure, through careful discussion and planning of the interventions, that the highly-experienced experimenter made both conditions equally appealing, motivating and interesting for children. Overall, therefore, despite the current study having some limitations, the findings hold promise for the utility of future short and enjoyable WM interventions that can enhance comprehension and working memory skills in young children with a relatively quick and easily implemented treatment.

The current findings suggest avenues for future research on far transfer effects that are relevant to underlying mechanisms of change (Melby-Lervåg & Hulme, 2012). Reading comprehension, language comprehension and ability to remember / follow instructions (e.g., Holmes et al., 2009; St Clair-Thompson et al., 2010) are all promising outcome measures that assess the child’s ability to divide attention between processing and ‘information storage updating’. Thus, a potential explanation for the mechanism underlying improvements via WM training is that such interventions improve the child’s ability to divide attention appropriately between processing and storage during tasks that require both skills concurrently, and / or they improve the child’s ability to choose and retain relevant information whilst suppressing other information not currently required. These skills are directly relevant for reading / listening comprehension (Hulme & Snowling, 2009), so if future research can link improvements in ELWM abilities to improvements in language and reading comprehension, this could advance theorising about potential mechanisms of change, and have implications for WM training interventions with children who have language and / or reading comprehension difficulties.

<u>Summary</u>

A relatively quick, face-to-face, adaptive working memory training intervention produced reliable gains in working memory skills in typically developing children. Children randomly allocated to the intervention condition showed significantly larger gains on two trained executive-loaded working memory tasks (Listening Recall, Odd One Out Span) and on two untrained working memory tasks (Word Recall, Counting Recall). These "near transfer" effects were still evident at a six month follow-up. "Far transfer" effects were much less in evidence: gains in single word reading, mathematics and spelling did not differ between the two groups over a period of 12 months. However, the trained group showed significantly higher reading comprehension scores than the control group at the 12 month follow-up, which indicates that improvements in the ability to divide attention between processing and storage may have benefits for reading comprehension.

References

Alloway, T.P., Bibile, V. & Lau, G. (2013). Computerized working memory training: Can it lead to gains in cognitive skills in students? *Computers in Human Behavior, 29*, 632-638. http://dx.doi.org/10.1016/j.chb.2012.10.023

Archibald, L. M. D., & Gathercole, S. E. (2006). Nonword Repetition: A Comparison of Tests. *Journal of Speech, Language & Hearing Research, 49*(5), 970-983.

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

Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V. & Camos, V. (2009). Working memory span development: A time-based resource-sharing model account. *Developmental Psychology, 45*(2), 477-490. DOI: 10.1037/a0014615

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. DOI:510.1111/j.1467-7687.2010.01022.x.

Burgoyne, K., Duff, F. J., Clarke, P. J., Buckley, S., Snowling, M. J., & Hulme, C. (2012). Efficacy of a reading and language intervention for children with Down syndrome: a randomized controlled trial. *Journal of Child Psychology and Psychiatry, 53*(10), 1044-1053. DOI: 1010.1111/j.1469-7610.2012.02557.x.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96*(1), 31-42. DOI: 10.1037/0022-0663.1096.1031.1031.

Case, R., Kurland, M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*, 386-404.

Chein, J.M. & Morrison, A.B. (2010). Expanding the mind's workspace: Training and transfer effects with a complex working memory task. *Psychonomic Bulletin and Review, 17* (2), 193-199. doi: 10.3758/PBR.17.2.193

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behaviour*, *19*, 450–456.

Ebbels, S. H. (2007). Teaching grammar to school-aged children with Specific Language Impairment using shape coding, *Child Language Teaching and Therapy, 23*, 67-93.

Elliott, C. D., Smith, P., & McCullouch, K. (1996). *British Ability Scales (BAS II)* (2nd ed.). Windsor: NFER-Nelson.

Ellis Weismer, S., Evans, J., & Hesketh, L. J. (1999). An examination of verbal working memory capacity in children with specific language impairment. *Journal of Speech, Language & Hearing Research, 42*(5), 1249-1260.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309-331. DOI: 310.1037/0096-3445.1128.1033.1309.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science (Wiley-Blackwell), 17*(2), 172-179. DOI:110.1111/j.1467-9280.2006.01681.x.

Gray, S.A., Chaban, P., Martinussen, R., Goldberg, R., Gotlieb, H., Kronitz, R., Hockenberry, M. & Tannock, R. (2012). Effects of a computerized working memory training program on working memory, attention, and academics in adolescents with severe LD and comorbid ADHD: A randomized controlled trial. *Journal of Child Psychology and Psychiatry, 53* (12), 1277-1284. doi: 10.1111/j.1469-7610.2012.02592.x.

Henry, L. A. (2001). How does the severity of a learning disability affect working memory performance? *Memory, 9,* 233-247.

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science, 12*(4), F9-F15. DOI:10.1111/j.1467-7687.2009.00848.x.

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,* 827-836. doi: 10.1002/acp.1589.

Hulme, C., & Snowling, M. J. (2009). *Developmental Disorders of Language, Learning and Cognition.* Chichester, West Sussex: Wiley-Blackwell.

Kail, R. (1990). *The Development of Memory in Children* (Third ed.). New York: W. H. Freeman.

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences, 14*(7), 317-324. DOI: 310.1016/j.tics.2010.1005.1002.

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.

Klingberg, T., Fernell, E., Olesen, P.J., Johnson, M., Gustafson, P., Dahlström, K. et al., (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. doi: 10.1097/00004583-200502000-00010

Leather, C. V., & Henry, L. A. (1994). Working memory span and phonological awareness tasks as predictors of early reading ability. *Journal of Experimental Child Psychology, 58*(1), 88-111. DOI: 110.1006/jecp.1994.1027.

Loosli, S. V., Buschkuehl, M., Perrig, W. J., & Jaeggi, S. M. (2012). Working memory training improves reading processes in typically developing children. *Child Neuropsychology, 18*(1), 62-78. DOI: 10.1080/09297049.09292011.09575772.

Martinussen, R., Hayden, J., Hogg-Johnson, S., & Tannock, R. (2005). A meta-analysis of working memory impairments in children with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 44*(4), 377-384. DOI: 310.1097/1001.chi.0000153228.0000172591.0000153273.

Melby-Lervåg, M., & Hulme, C. (2012). Is working memory training effective? A meta-analytic review. *Developmental Psychology, 49* (2), 270–291. , DOI: 10.1037/a0028228.

Nation, K., Adams, J. W., Bowyer-Crane, C. A., & Snowling, M. J. (1999). Working Memory Deficits in Poor Comprehenders Reflect Underlying Language Impairments. *Journal of Experimental Child Psychology, 73*(2), 139-158.

Nutley, S.B., 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. doi: 10.1111/j.1467-7687.2010.01022.x.

Passolunghi, M. C. (2006). Working memory and arithmetic learning disability. In T. P. Alloway & S. E. Gathercole (Eds.), *Working memory and neurodevelopmental disorders* (pp. 113-138). Hove, UK: Psychology Press.

Pickering, S., & Gathercole, S. E. (2001). *Working Memory Test Battery for Children (WMTB-C).* London: The Psychological Corporation Ltd.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *138*(4), 628-654. doi: 610.1037/a0027473.

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. DOI: 210.1080/01443410903509259.

Thorell, L.B., Lindqvist, S., Nutley, S.B., Bohlin, G. & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science, 12*:1, 106-113. doi: 10.1111/j.1467-7687.2008.00745.x.

Swanson, H. L. (2006). Working memory and reading disabilities: Both phonological and executive processing deficits are important. In T. P. Alloway & S. E. Gathercole (Eds.), *Working memory and neurodevelopmental disorders* (pp. 59-88). Hove, UK: Psychology Press.

Van der Molen, M. J., Van Luit, J. E. H., Van der Molen, M. W., Klugkist, I., & Jongmans, M. J. (2010). Effectiveness of a computerised working memory training in adolescents with mild to borderline intellectual disabilities. *Journal of Intellectual Disability Research*, *54*(5), 433-447. DOI:410.1111/j.1365-2788.2010.01285.x.

Wechsler, D. (1993). *Wechsler Objective Reading Dimensions*. London: The Psychological Corporation.

Table 1. Mean scores on all study measures (raw and standard scores) at all relevant time points. Matrices and Verbal Similarities standard scores are expressed as BAS-II T-scores (mean 50, SD 10), working memory raw scores are total trials correct, and scholastic skills raw scores are given as 'ability' scores (for the BAS II) or raw scores (for the WORD).

		Intervention group			
		Training Intervention		Treated Control	
Test	Test point	M (SD) Raw scores	M (SD) Standard scores	M (SD) Raw scores	M (SD) Standard scores
Single-word reading (BAS-II)	Pre	106.22 (24.72)	113.28 (13.56)	98.56 (24.46)	109.69 (14.04)
	Post	111.83 (23.25)	113.50 (15.11)	101.81 (25.43)	108.63 (14.54)
	6m	121.33 (22.58)	114.11 (15.96)	109.94 (20.75)	108.31 (13.84)
	12m	126.44 (21.97)	112.22 (15.39)	116.31 (17.82)	106.88 (14.12)
Number skills (BAS-II)	Pre	71.11 (20.57)	106.50 (7.32)	69.31 (19.52)	105.88 (8.71)
	Post	79.56 (17.40)	109.89 (10.52)	76.81 (17.74)	109.06 (10.76)
	6m	85.89 (14.44)	109.89 (11.26)	82.25 (15.93)	108.13 (11.05)
	12m	87.78 (15.08)	106.17 (7.71)	86.75 (14.53)	105.25 (13.12)
Spelling (BAS-II)	12m	98.06 (20.72)	115.06 (17.32)	89.44 (13.78)	109.0 (14.98)
Reading comprehension (WORD)	12m	20.33 (4.34)	109.28 (9.11)	15.81 (4.92)	98.44 (11.71)
Digit Recall (WMTB-C)	Pre	27.28 (3.77)	107.61 (13.87)	26.88 (3.08)	106.00 (10.00)
	Post	29.50 (4.18)	115.06 (15.71)	27.65 (3.84)	109.11 (12.71)
	6m	30.78 (4.75)	118.50 (18.40)	29.18 (3.34)	112.71 (14.03)

Word Recall (WMTB-C)	Pre Post 6m	18.11 (2.19) 21.94 (3.47) 22.78 (2.37)	103.61 (12.63) 120.17 (16.95) 120.61 (13.32)	16.88 (1.90) 18.35 (2.60) 18.41 (2.09)	98.72 (10.09) 105.44 (9.89) 100.76 (9.23)
Block Recall (WMTB-C)	Pre Post 6m	21.89 (3.63) 23.50 (3.48) 23.39 (3.13)	97.28 (10.72) 102.67 (10.71) 98.89 (11.30)	22.12 (2.76) 22.88 (2.50) 23.00 (3.92)	98.00 (15.20) 101.50 (15.60) 98.12 (16.80)
Listening Recall (WMTB-C)	Pre Post 6m	9.33 (2.54) 15.22 (3.59) 14.50 (2.71)	100.22 (12.25) 129.67 (16.81) 123.17 (13.56)	7.82 (2.24) 8.88 (2.85) 9.29 (2.09)	93.94 (10.78) 98.67 (12.56) 96.24 (12.04)
Counting Recall (WMTB-C)	Pre Post 6m	14.11 (4.09) 18.72 (3.75) 18.89 (3.22)	88.67 (9.88) 102.17 (11.35) 99.72 (10.91)	14.29 (4.54) 14.82 (3.91) 16.24 (3.65)	91.11 (12.02) 91.44 (12.08) 89.71 (13.98)
Odd One Out Span	Pre Post 6m	5.89 (2.06) 9.78 (3.00) 10.50 (2.98)	- - -	5.18 (2.10) 6.18 (1.91) 6.65 (2.15)	- - -

Captions

Figure 1: Tasks administered at each of the four time points during the study

Figure 2: Graphs showing training gains for the WM training intervention and control group on all study measures (with standard error bars).

Figure 1

Pre-Test
BAS II Matrices
BAS II Verbal Similarities
BAS II Number Skills
BAS II Word Reading
WMTB-C Digit Recall
WMTB-C Word Recall
WMTB-C Block Recall
WMTB-C Listening Recall
WMTB-C Counting Recall
Odd One Out Span

Post-Test
BAS II Number Skills
BAS II Word Reading
WMTB-C Digit Recall
WMTB-C Word Recall
WMTB-C Block Recall
WMTB-C Listening Recall
WMTB-C Counting Recall
Odd One Out Span

6 Month Follow-up
BAS II Number Skills
BAS II Word Reading
WMTB-C Digit Recall
WMTB-C Word Recall
WMTB-C Block Recall
WMTB-C Listening Recall
WMTB-C Counting Recall
Odd One Out Span

12 Month Follow-up
BAS II Number Skills
BAS II Word Reading
BAS II Spelling
WORD Reading Comprehension

Figure 2

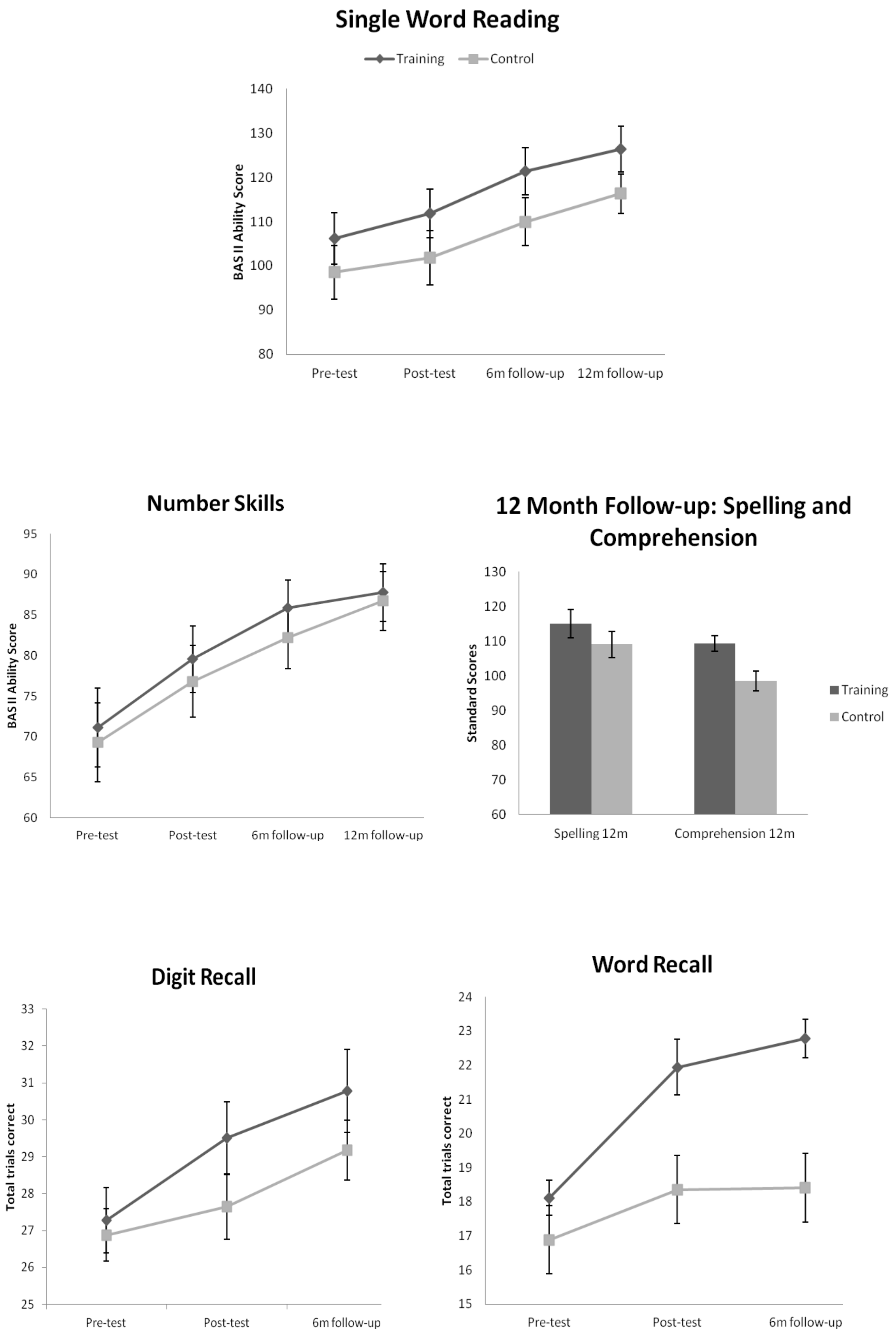
Single Word Reading
Training
Control
BAS II Ability Score
140
130
120
110
100
90
80
Pre-test
Post-test
6m follow-up
12m follow-up
Number Skills
BAS II Ability Score
95
90
85
80
75
70
65
60
Pre-test
Post-test
6m follow-up
12m follow-up
12 Month Follow-up: Spelling and Comprehension
Standard Scores
130
120
110
100
90
80
70
60
Spelling 12m
Comprehension 12m
Training
Control
Digit Recall
Total trials correct
33
32
31
30
29
28
27
26
25
Pre-test
Post-test
6m follow-up
Word Recall
Total trials correct
24
23
22
21
20
19
18
17
16
15
Pre-test
Post-test
6m follow-up

Figure 2 continued