

Working Memory Abilities and Children's Performance in Laboratory Analogues of Classroom Activities

SUSAN E. GATHERCOLE^{1*}, EMILY DURLING², MATTHEW EVANS²,
SARAH JEFFCOCK² and SARAH STONE²

¹*University of Durham, Durham, UK*

²*University of York, York, UK*

SUMMARY

Laboratory analogues of classroom activities on which children with low working memory skills have been observed to perform very poorly were developed and employed in two studies. In Study 1, 5- and 6-year-old completed one task involving recalling spoken sentences and counting the numbers of words, and another task involving the identification of rhyming words in spoken poems. Poorer performance of low than average working memory children was obtained on the recall measure of both tasks. In Study 2, 5- and 6-year-old children heard spoken instructions involving the manipulation of a sequence of objects, and were asked either to perform the instructions or repeat them, in different conditions. The accuracy of performing but not repeating instructions was strongly associated with working memory skills. These results indicate that working memory plays a significant role in typical classroom activities that involve both the storage and mental manipulation of information. Copyright © 2007 John Wiley & Sons, Ltd.

One of the principal factors associated with a child's ability to learn is working memory, the capacity to hold in mind and manipulate information for brief periods of time in order to complete an ongoing mental activity. Progress in the key academic domains of reading and mathematics is closely related to children's working memory skills (Gathercole, Pickering, Knight, & Stegmann, 2004; Geary, Hoard, Byrd-Craven, & DeSoto, 2004; Swanson, Ashbaker, & Lee, 1996). Recent evidence also indicates that the majority of children identified with reading difficulties have very poor working memory function and that within this population, the severity of reading and mathematics difficulties is more closely associated with working memory capacity rather than other related verbal abilities such as IQ and phonological awareness (Gathercole, Alloway, Willis, & Adams, 2006; see also, Siegel & Ryan, 1989).

Exactly why the ability to learn is so closely related to working memory is not well understood, and to date has been the subject of little direct investigation. One proposal is that working memory plays an important role in the integration of current inputs with information retrieved from long-term memory, and that this process is compromised by poor working memory capacity (Swanson & Saez, 2003; Swanson & Beebe-Frankenberger, 2004). A related suggestion is that many of the individual classroom learning activities that

*Correspondence to: Professor Susan E. Gathercole, Department of Psychology, University of York, Heslington, York, YO10 5DD, UK. E-mail: s.gathercole@psychology.york.ac.uk

form the basis for the child's learning place heavy demands on working memory; children with poor working memory capacity may fail to meet these demands, leading to forgetting of task-relevant information and ultimately, task failure (Gathercole, Lamont, & Alloway, 2006). These task failures represent missed learning opportunities that will slow the cumulative rate of the acquisition of skills and knowledge, resulting in slow rates of learning and poor academic progress.

The most direct way to investigate the constraints that working memory abilities may impose on learning is to study children at the point at which most formal learning takes place: in the classroom. Preliminary support for the working memory overload hypothesis was provided by classroom observation of 5- and 6-year-old children who had scored very poorly on measures of working memory at school entry 1 year earlier (Gathercole et al., 2006). Working memory ability was assessed using complex memory span tests such as listening span in which the participant must both process and store new information (Daneman & Carpenter, 1983). A year later, these children were in the lowest ability groups in reading and mathematics. They were also observed to make frequent failures in three types of activities that impose high working memory loads: remembering the content of classroom instructions (see also, Engle, Carullo, & Collins, 1991), performing activities that require both the storage and processing of material, and keeping track in complex tasks. In these situations, the children typically started the activity, started to make errors and failed to see it through to completion.

The purpose of the present research was to evaluate the impact of poor working memory on some of these learning activities when they are performed under more controlled conditions. In two studies, children with low and average working memory scores completed laboratory analogues of these tasks. In Study 1, one task mimicked a classroom activity in which children were asked to count the words in a dictated sentence before writing the sentence down (Gathercole et al., 2006). In the laboratory version, the children were requested to count the words in spoken sentence and then to recall the sentence. In this activity, it is necessary to maintain an accurate representation of the sentence in working memory for sufficiently long to count the individual words and then to guide the recall attempt. The concurrent processing element of the task is to count the words, a taxing activity for 5- and 6-year-old children involving both the identification of individual lexical items in the sentence and the use of number knowledge to count the words. The accuracy of both counting and recall of the sentence was recorded. The counting measure provides an estimate of processing ability, allowing us to assess whether any deficits found in this task with the low memory children in sentence recall are mediated by the processing efficiency. Although recent evidence indicates that individual variation in working memory typically cannot be accounted for simply by differences in processing (Bayliss, Jarrold, Gunn, & Baddeley, 2003; Saito & Miyake, 2004), it was important to investigate the potential impact of processing on recall in this new paradigm.

The second classroom analogue task employed in this study was based on a classroom activity in which a teacher read aloud a poem and asked children to identify the two rhyming words it contained (Gathercole et al., 2006). This task imposes a significant working memory burden, as the child must maintain the spoken material for sufficiently long to make the necessary pair-wise phonological comparisons of words required to identify the rhyming items. In order to investigate the possible impact of the processing element of the task on the recall of the rhyming word pair, the children were also assessed on their ability to detect pairs of rhymes using a standardised measure from the Phonological Abilities Test (Muter, Hulme, & Snowling, 1997).

A further issue addressed in this study concerns whether attentional difficulties accompany poor working memory. Teachers of the low memory children observed by Gathercole et al. (2006) often described them as being inattentive, with short attention spans. Other evidence also points to a link between working memory and inattentive behaviour. Martinussen and Tannock (2006) reported that inattentive symptoms in a sample of children with ADHD were closely associated with working memory skills, and Aronen, Vuontela, Steenari, Salmi, and Carlson (2005) found a similar association in a sample of typically developing children. More directly, Klingberg et al. (2005) reported that a computerised working memory training programme administered to children with ADHD not only boosted their working memory scores, but also led to improvements in parent ratings of inattentive behaviours. The present studies provided the opportunity to establish directly whether inattentive behaviour also characterised the children with poor working memory. To this end, teachers completed the Conners' Teacher Rating Scale-Revised (2001), a behaviour ratings measure designed to identify symptoms associated with ADHD and other attentional problems.

STUDY 1

Method

Participants

Participants were identified from a sample of 100 5- and 6-year-old children attending two local education authority primary schools with mixed intakes in an urban area of the north of England. Written consent was gained from parents or guardians prior to participation. In the screening phase, children were tested individually in a quiet environment on two measures of verbal complex memory from the Automated Working Memory Assessment (AWMA, Alloway, 2007; see also, Alloway, Gathercole, & Pickering, 2006). Both tests were administered on a laptop PC. The listening recall test involves the child listening to spoken sentences, judging whether each sentence is true and false and attempting to recall the final word of each of the sentences presented on a trial in their original sequence. In the backward digit recall test, the child attempts to recall a sequence of auditorily presented digit names in their reverse serial order. In both cases, standard scores (population mean = 100, SD = 15) were computed and a mean working memory score was calculated from the two test scores. Test-retest coefficients were .81 for listening recall, and .64 for backward digit recall.

Two groups of children were identified on the basis of mean working memory scores: the low working memory group consisted of 15 children with scores equal to or less than 85 formed the low working memory group, and the average working memory group consisted of 15 children with working memory scores between 95 and 105. There were seven females and eight males in each group. Descriptive statistics for the working memory scores are provided in Table 1.

Procedure

Each member of the low and average working memory groups participated in a further individual testing session located in a quiet area of the child's school. The following measures were administered during this session. Two experimenters attended each session,

Table 1. Descriptive statistics for cognitive and behavioural measures

Measure	Working memory group			
	Low		Average	
	<i>M</i>	SD	<i>M</i>	SD
Working memory				
Listening recall	73.80	4.81	102.27	8.22
Counting recall	82.60	8.43	95.27	7.72
Composite	78.20	5.48	98.77	3.75
Rhyme oddity detection	7.07	2.89	9.00	2.20
Raven's Matrices	14.80	3.14	17.47	4.17
Teacher Rating Scale				
Oppositional	48.13	10.87	47.17	6.79
Cognitive problems/inattention	64.50	19.05	48.50	6.06
Hyperactivity	50.63	8.60	48.67	8.64
ADHD index	52.75	8.46	47.00	4.82

which was also recorded onto a laptop PC. Scores were recorded manually at the time of testing by one experimenter.

Classroom analogue tasks

Sentence counting and recall. Presentation of this task was controlled by the PC, using digitised audio files recorded for each sentence. Instructions were presented auditorily, and the child was asked if s/he understood the task before commencing the practice trials. On each trial, a simple sentence such as *Beth likes to paint pretty pictures* was spoken. The child's task was to count the number of words in the sentence, stating the total aloud, and then to repeat the sentence (the materials are provided in Appendix 1). Following four practice trials with feedback, the child received 20 experimental trials consisting of three sentences each containing 2, 3, 4, 5 and 6 words, two sentences each containing 7 and 8 words and one sentence containing 9 words. All children received the sentences in the same sequence, in which sentence length varied unsystematically across trials. The accuracy of the counting and recall attempt on each trial was scored, and the total number of sentences correctly counted and correctly recalled was calculated for each child.

Detecting rhymes in poems. Two-line extracts from poems taken from educational materials devised for this age group of children were employed in this task (see Appendix 2 for stimuli). A PC controlled presentation of the instructions and stimuli, and recorded the spoken responses. The children received two practice trials with feedback in which the tester read aloud a poem, and asked the child to say the two rhyming words. Seven test trials were then presented without feedback. The accuracy of responses was recorded, and the number of word pairs correctly identified was scored for each child.

Standardised measures

Rhyme detection. The rhyme detection test from the Phonological Abilities Test (Muter et al., 1997) was administered. This test consists of three practice trials followed by 10 test trials. On each trial the child is shown a card displaying drawings of four familiar objects

with one-syllable labels, which are named by the experimenter. The child is asked to identify the picture whose name rhymes with that of the first picture (e.g. *key* and *tree*). The number of trials on which a correct response is made is calculated.

Non-verbal reasoning. The Progressive Matrices test (Raven, 1997) was administered to each child. On each trial, the child views a two-dimensional pattern with a missing segment, and six individual segments. The task is to identify the segment that completes the pattern. A total of 36 patterns are viewed, and the total number of correct responses is scored.

Teacher behaviour ratings. The Teacher Rating Scale-Revised Short Form (Conners, 2001) was completed by the classroom teacher of each participating child. This measure is designed to identify attentional failures and ADHD on the basis of classroom behaviours, and involves teachers rating the extent to which the child has had problem behaviours in school over the past month that are described in 28 brief statements on the form. Responses are scored as sums of values on four subscales—oppositional (e.g. spiteful or vindictive), cognitive problems/inattention (e.g. forgets things s/he has already learned), hyperactivity (e.g. is always 'on the go' or acts as if driven by a motor) and ADHD index (e.g. restless, always up and on the go). The ADHD Index is based on behaviours closely associated with a clinical diagnosis of ADHD. *T* scores (with a population mean of 50 and SD of 10) are calculated for each of the four subscales.

Results

Standardised measures

Descriptive statistics for scores on each of the standardised measures are provided in Table 1. A MANOVA performed on the two working memory test scores established a highly significant effect of group by Hotelling's *T*: $F(2,27) = 87.21$, $p < .001$, partial $\eta^2 = .866$. The outcomes of univariate *F* tests performed on the scores as a function of working memory were as follows: rhyme detection, $F(1,28) = 4.244$, $p = .049$, partial $\eta^2 = .132$; Progressive Matrices, $F(1,28) = 3.908$, $p = .058$, partial $\eta^2 = .122$.

In a MANOVA performed on the Teacher Rating Scale subscale scores, the effect of group was highly significant: Hotelling's *T*, $F(4,25) = 7.14$, $p = .001$, partial $\eta^2 = .990$. Univariate *F* tests on the subscale scores established significant effects of group on two measures: cognitive problems/inattention, $F(1,28) = 25.084$, $p < .001$, partial $\eta^2 = .473$ and ADHD index, $F(1,28) = 6.967$, $p < .05$, partial $\eta^2 = .199$; for the remaining two measures, $p > .10$. According to the interpretive guidelines for the test (Conners, 2001), *T*-scores of 55 or below do not represent a cause for concern, scores in the range 56–60 are slightly atypical and should raise concern, scores of 61–65 are mildly atypical and represent a possible significant problem, scores of 66–70 are moderately atypical and represent a significant problem and scores 70 and greater are markedly atypical. The majority of children in both groups obtained scores of 55 and below on the oppositional subscale (93% of both groups) and the hyperactivity subscale (87% of the low memory group and 93% of the average memory group). In contrast, only 20% of the low memory group obtained scores in this typical range on the cognitive problems/inattention subscale with 40% scoring greater than 70 (markedly atypical); the corresponding figures for the average memory group were 93% and 0%, respectively. On the ADHD index subscale, scores of 55 and below were obtained by 60% of the low memory group, and by 93% of the

Table 2. Descriptive statistics for the classroom analogue tasks, Study 1

Task	Working memory group			
	Low		Average	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Sentence counting and recall				
Counting	7.07	5.66	11.73	5.57
Recall	9.73	4.64	15.07	4.01
Rhyme detection in poems	1.40	1.59	4.73	2.40

average memory group. No child in either group obtained scores in excess of 70. Thus, the most marked differences between groups were found for the behaviour problems relating to cognitive problems and inattention, for which the majority of the low memory group obtained elevated scores in the atypical range.

Sentence counting and recall

Descriptive statistics for the measures from the sentence counting task are provided in Table 2. In both cases, the average working memory group obtained significant higher scores than the low working memory group. Significant group effects were found in ANOVAs performed on both dependent variables: counting, $F(1,28) = 5.174$, $p = .031$, partial $\eta^2 = .156$; recall, $F(1,28) = 11.359$, $p = .002$, partial $\eta^2 = .289$. In order to determine the possible mediating effects of group differences on test scores in general cognitive ability, an ANCOVA was conducted for each dependent variable in which non-verbal ability scores (Progressive Matrices) were entered as a covariate. The interaction between group and the covariate was non-significant in both ($p > .5$). The group effect remained significant on the recall measure, $F(1,27) = 8.242$, $p = .008$, partial $\eta^2 = .234$, but not on counting, $F(1,27) = 2.182$, $p = .151$, partial $\eta^2 = .075$. In a further ANCOVA designed to establish whether group differences in recall could be attributable to counting ability, both Progressive Matrices and counting scores were entered as covariates: the group effect remained significant, $F(1,26) = 5.402$, $p = .028$, partial $\eta^2 = .172$; for non-verbal reasoning, $F(1,26) < 1$, partial $\eta^2 = .007$ and for counting, $F(1,26) = 8.593$, partial $\eta^2 = .248$.

The counting and recall data were further analysed as a function of the number of words in the sentences. Accuracy of both counting and recall at each sentence length is summarised in Figures 1A,B, from which it can be seen that performance on both measures declines with increasing length.

Detecting rhymes in poems

It can be seen from Table 2 that the low working memory group performed more poorly on this task than the average working memory group. The distribution of scores differed markedly across the two groups, as shown in Figure 2: 73% of the low memory group but only 13% of the average memory group obtained scores of either 0 or 1. The group effect was highly significant in an ANOVA performed on the test scores, $F(1,28) = 20.023$, $p < .001$, partial $\eta^2 = .417$. In order to test the specificity of this group effect, an ANCOVA was performed in which non-verbal reasoning (Progressives Matrices) and rhyme detection (Phonological Abilities Test) scores were entered as covariates. In this analysis, the group effect remained, $F(1,26) = 10.124$, $p = .004$, partial $\eta^2 = .280$; for non-verbal reasoning,

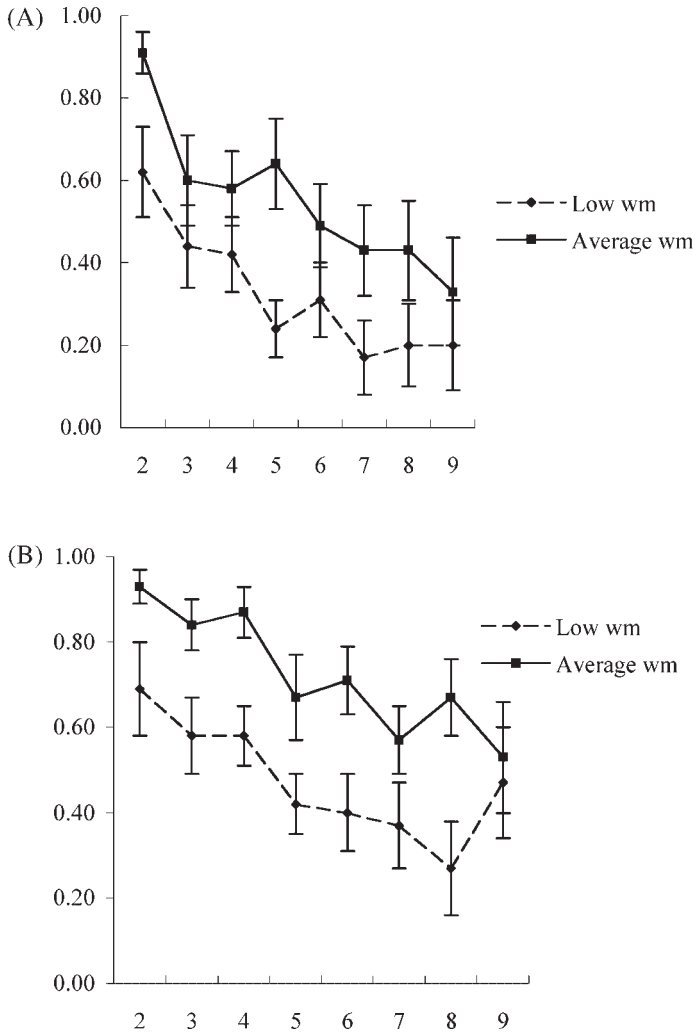


Figure 1. Mean scores (with SE bars) as a function of number of words in sentence and group in Study 1: accuracy of counting (A) and recall (B)

$F(1,26) = 7.304, p < .05, \text{partial } \eta^2 = .219$ and for rhyme detection, $F(1,26) = 3.059, p > .05, \text{partial } \eta^2 = .105$. In a further ANCOVA in which interactions between group and each of the covariates were included as terms in the model, both interaction terms were non-significant ($p > .10$, in both cases).

Discussion

In both tasks—counting the words in a spoken sentence and then attempting to recall the sentence, and identifying rhyming words in spoken two-line poems—poorer performance was achieved by the low than the average working memory group. These findings substantiate previous observations that children with low verbal working memory scores

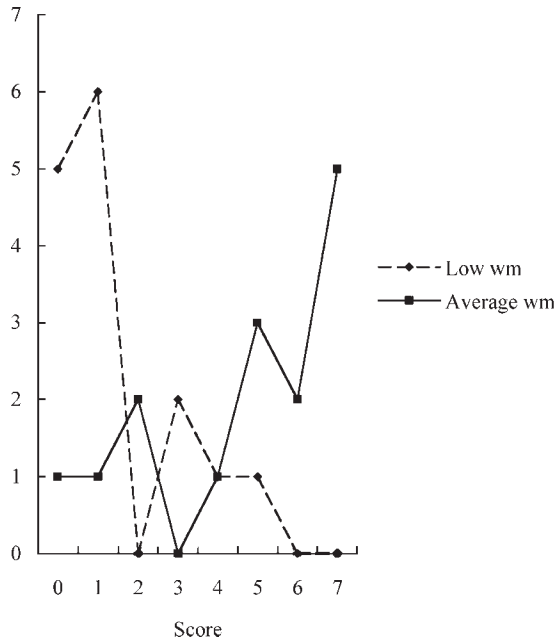


Figure 2. Frequency of scores (accuracy of detecting rhyme pairs) as a function of group in Study 1

have marked difficulties in these classroom activities, both of which impose substantial demands on temporary storage and processing (Gathercole et al., 2006).

In line with previous research employing more conventional assessments of working memory (Bayliss et al., 2003; Saito & Miyake, 2004), individual differences in the recall performance in both tasks could not be accounted for solely in terms of general cognitive abilities or of skills in the processing element, although these factors did have some influence on scores. Thus, sentence recall remained significantly lower in the low memory group even when differences in non-verbal and counting ability were taken into account, and their relative deficit in identifying rhymes in poems persisted when scores were adjusted for both non-verbal ability and for abilities to detect rhymes in a task in which memory demands were minimised.

A distinctive pattern of classroom problem behaviours relating to attention was established for this sample of low working memory children on the Conners' Teacher Rating Scale (2001). The majority of the children were rated to have elevated levels of behaviours indicative of cognitive problems (specifically, to making poor academic progress in literacy and mathematics) and inattention; the group also scored more highly than the average memory group on the ADHD index subscale, although the difference was not as marked as for the cognitive problems/inattention subscale. The same pattern of rated problem behaviours was also recently found in a larger group of children aged 5–6 years and 9–10 years selected on the basis of low working memory scores (Gathercole, Alloway, Kirkwood, Elliott, Holme, & Hilton, in press). Low rates of academic progress, inattentivity and poor working memory therefore appear to co-occur.

Problems in meeting the working memory demands of common classroom-based activities in another common classroom situation—following verbal instructions to

perform actions—were investigated in Study 2. Gathercole et al. (2006) observed that low working memory children often failed to complete complex but by no means unusual classroom instructions such as *Put your sheets on the green table, arrow cards in the packet, put your pencil away and come and sit on the carpet*; typically, they performed only the first one or two commands of the sequence before giving up. The storage load on working memory in this situation is significant, requiring the maintenance of a representation of the sequence across the protracted period during which the individual actions of the sequence are carried out. The child must also be able to keep track of his or her place in the task, an ability that also appears to be compromised in individuals with low working memory (Gathercole et al., 2006). Preliminary evidence that working memory is directly associated with the ability to follow instructions was provided by Engle et al. (1991). They employed a following directions task in which children aged 6–12 years attempted to follow directions based on such as *Sit on the floor Indian-style* and *Point to the picture at the top of page three and copy it twice* that varied systematically in the number of actions and quantifiers. Importantly, performance accuracy was found to be significantly associated with measures of working memory ability.

Study 2 investigated this link by employing a task in which children either followed instructions to manipulate sequences of objects placed in front of them, or repeat the instructions. The length of the instruction sequence increased to the point at which the child could not longer perform the task accurately. The extent to which the children's working memory scores were related to their abilities to perform the action sequences and to repeat the instructions was the main focus of this study. The instruction paradigm used here shares many features with the Token Test (De Renzi & Faglioni, 1978), a measure developed for use in neuropsychological assessment in which sequences of commands are given that participants must perform by manipulating objects on a display. Across blocks, the length and complexity of the instructions increases, and patients with acquired deficits of verbal short-term memory typically have difficulties with lengthy and complex instructions (Bartha & Benke, 2003).

Our instructions task had two key features that distinguish it from the Token Test. First, the instructions varied in the number of steps but not in the complexity of the grammatical constructions: only simple active constructions were used, concatenated with the adverb *then*. Second, a span-type procedure was employed in which the length of the instruction sequence increased systematically up to the point at which the child could not longer perform the task with reasonable accuracy. Action sequences were only scored as correct if they reflected the serial order of the instructions. The instructions required to child to perform simple actions on an array of objects and containers, such as *Touch the white bag then pick up the yellow ruler then put it in the blue folder*. The number of steps in the instructions increased progressively up to the point at which the child was no longer able to perform accurately.

In a second condition, the children repeated the instructions rather than performing them. It was anticipated that performance would place greater demands on working memory than repetition, both as a consequence of the longer interval over which the instruction must be maintained in working memory, and of the necessity to keeping track of which actions had been executed and which had not. Our observational study of the classroom functioning indicated that pace-keeping difficulties were commonplace in children with low working memory (Gathercole et al., 2006).

As in Study 1, teacher ratings of attentional problem behaviours (Conners, 2001) were also obtained for the children. This enabled us to investigate whether the atypically high

levels of problem behaviours relating to cognitive problems and inattention found in the low memory group in Study 1 would be replicated in an independent sample.

STUDY 2

Method

Participants

One hundred children (54 boys, 46 girls) attending reception classes of eight primary schools in the North East of England participated in the standardisation of the AWMA (Alloway, 2007). The mean age of the sample was 5 years 1 month (SD = 4.031 months, min = 4 years 6 months, max = 5 years 9 months). Schools were selected to approximate closely to national standards in Key Stage 2 national curriculum assessments. Two subgroups of children from this sample were identified for the purposes of the present study on the basis of their scores of the three verbal working memory subtests of the AWMA: listening recall, backwards digit recall and counting recall. The listening recall test involves the child listening to spoken sentences, judging whether each sentence is true and false and attempting to recall the final word of each of the sentences presented on a trial in their original sequence. In the backward digit recall test, the child attempts to recall a sequence of auditorily presented digit names in their reverse serial order. In counting recall, the child counts aloud the number in a number of successive arrays, and then attempts to recall the totals in the original sequence. In each test, standard scores (population mean = 100, SD = 15) were computed, and a mean working memory score was calculated from the three test scores. Test retest reliabilities for a sample of children aged 4.5–11.5 years were .81, .79 and .64 for listening recall, counting recall and backwards digit recall, respectively.

The present sub-sample of children was composed of 15 children (8 girls and 7 boys, mean age 5 years 9 months) with a mean working memory score (based on the three standard scores) of 88 or less, and 14 children (4 girls and 10 boys, mean age 5 years 11 months) who obtained scores in the range 97–104. The mean score of the low working memory group was 84.62; for the average working memory group, the mean score was 100.47. On average, they participated in the present study approximately nine months after completing the AWMA. In order to provide an assessment of working memory at the time of testing in the present study, forwards and backwards digit recall were tested approximately 2 months after completion of the instructions test. The Progressive Matrices Test on non-verbal reasoning (Raven, 1997) was also administered at this time.

Design and materials

The instructions consisted of descriptions of actions to be performed on a subset of six objects (a yellow ruler, a blue ruler, a white eraser, a green eraser, a red pencil and a black pencil) and three containers (a black box, a red box, a yellow bag, a white bag, a blue folder and a green folder). The individual actions in the instructions involved touching (e.g. *touch the red pencil*) and picking up (e.g. *pick up the yellow ruler and put it in the black box*). The items (objects and containers) used in each instruction were selected at random, with the constraints that there was no repetition of a colour, object or contained within an individual action, or of a colour/object or colour/container combination in the instruction as a whole. Actions involving touching and pickup up were concatenated using the adverb *then* in order

to produce instructions involving different numbers of items. Thus, an example of a five-item sequence is *Touch the red pencil THEN Pick up the blue ruler and put it in the black box THEN Pick up the white eraser and put it in the red box.*

Two conditions were tested for each participant, in which the instructions were either repeated or enacted. The order of testing the two conditions was randomised. In each condition, test trials were organised into blocks of three trials containing instructions involving 1–8 items.

Procedure

Each child was tested individually in a quiet area of the school, seated at a table opposite the test administrator. Objects and containers were laid out on a table within arm's reach of the child. All containers were closed to ensure the child could not see what they had placed inside. Before testing commenced, each item was named and the child was asked to point to it, in order to familiarise the child with the item's label and its location. The instructions were read out to the child at a measured rate. The child was asked either to repeat the instruction back in the repetition condition, or to do the things the test administrator said in the action condition.

Test trials in both the repetition and action conditions began with a block of one-item instruction and increased by one action per block until the point at which the child failed to complete two trials in a block correctly, at which point testing was stopped. Performance was scored according to a strict serial recall criterion, in which an action had to be repeated or performed in its correct position within the instruction to be scored as correct. Four measures of accuracy were scored in each trial: the number of correct features (colour, object or container), the number of correct items (colour/object and colour/contained combinations) and the number of correct instructions (all actions correct). These scores were summed across trials to produce a total recall score for features, items and sequences in each condition. In addition, a measure of span which was the maximum number was computed in each condition, which was the maximum number of items on which a participant performed correctly on at least two trials, plus a credit of .33 for a further correct trial at the next length if appropriate.

Results

Descriptive statistics for the cognitive assessments and the behaviour rating measure are provided in Table 3. For the sample as a whole, standard scores for backwards digit recall were in the low average range and for digit recall in the high average range. Average *T* scores on the Teacher Rating Scale fell in the typical range of 55 and below for three of the subscales—oppositional, hyperactivity and ADHD index—but were elevated into the atypical range for the cognitive problems/inattention subscale ($M = 60.41$), with a relatively large standard deviation.

Scores on the two conditions of the instructions task—action and repetition—are summarised in Table 4. One substantial and unexpected feature of the data was that performance was more than twice as accurate in the action than the repetition conditions, for each of the different measures – features, items, sequences, and span.

In order to explore possible associations between memory scores and performance on the instructions task, a correlation matrix was computed that included all action and repetition scores, the two memory measures obtained shortly after the instructions task was completed (forward and backward digit recall), non-verbal ability scores and subscale

Table 3. Descriptive statistics for cognitive and behavioural measures, Study 2

Measure	<i>M</i>	SD	Kurtosis	Skewness
Backwards digit recall	93.86	15.27	0.01	-0.38
Digit recall	106.00	12.62	0.22	-0.59
Raven's Matrices	19.90	5.11	0.66	0.15
Teacher Rating Scale				
Oppositional	47.79	8.60	4.27	20.07
Cognitive problems/inattention	60.41	18.16	0.40	-1.61
Hyperactivity	49.62	8.15	2.17	6.07
ADHD index	54.10	12.73	1.34	1.15

scores on the Teacher Rating Scale. The backward digit recall measure was correlated to a moderate though significant level with the same measure administered 11 months earlier ($r = .331, p = .039$ by one-tailed test, with 27 df). The two memory measures approximated reasonably well to a normal distribution (see kurtosis and skewness values in Table 3). However, two other measures had distributional properties that deviated substantially from a normal distribution—the oppositional and hyperactivity subscales from the Teacher Rating Scale. In both cases, the high kurtosis and skewness values reflected the far greater proportion of low than high scores, combined with bunching of scores due to a truncated distribution at the low end of the distribution, with the lowest obtainable *T*-score of 43 (with a population mean of 50). Of the 29 participants, only 2 had scores in the atypical range (greater than 55) on the oppositional scale, and 4 in the atypical range for the hyperactivity scale. Thus, only a minority of the participants were rated as having problems in these areas of behaviour. In contrast, the cognitive problems/inattention and ADHD index subscale scores had distributions that were close to normal. All subscale scores were included in the correlation matrix but due to the deviations from normality noted above for two of the scores, outcomes should be interpreted with caution.

The matrices of zero-order and partial correlation coefficients (controlling for Raven scores) are shown in Table 5. Working memory scores were highly and significantly associated with recall accuracy in the action condition, but not in the repetition condition. Moreover, the correlations with action recall were generally higher with backward than forward recall, particularly in the feature and item recall measures that credit partially correct recall responses. In order to assess the possible specificity of the associations of the two different working memory indices with recall in the action condition, partial correlation coefficients were computed. The correlations between backward digit recall and the action recall measures remained highly significant when digit recall scores were partialled out: with 26 df, $r = .523, p < .005$, for feature recall, $r = .537, p < .005$, for item recall, $r = .473, p < .05$, for sequence recall and $r = .477, p < .05$, for span. Conversely,

Table 4. Descriptive statistics for the action and repetition conditions

Measure	Action		Repetition	
	<i>M</i>	SD	<i>M</i>	SD
Feature recall	79.00	24.01	29.17	14.62
Item recall	37.55	12.00	13.24	7.09
Sequence recall	10.34	2.16	5.17	1.49
Span	3.75	0.74	1.85	0.37

Table 5. Correlations between principal measures: zero-order coefficients in lower triangle, coefficients with Raven scores partialled out in upper triangle

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1 Non-verbal reasoning	—														
2 Digit recall	.009	—													
3 Backwards digit recall	.351	.156	—												
4 Action, features	.298	.301	.540**	—											
5 Action, items	.343	.348	.551**	.980**	—										
6 Action, sequences	.360	.482*	.485**	.791**	.796**	—									
7 Action, span	.275	.508*	.485**	.790**	.801**	.924**	—								
8 Repetition, features	.160	.049	.172	.108	.129	.206	.206	—							
9 Repetition, items	.016	.071	.131	.103	.118	.193	.209	.936**	—						
10 Repetition, sequences	.073	.179	-.015	.002	.030	.136	.181	.527**	.547**	—					
11 Repetition, span	-.013	.058	-.093	-.004	.062	.169	.289	.499**	.539**	.625**	—				
12 Oppositional ^a	.008	.117	-.373*	-.209	-.186	-.129	-.155	.064	.075	.148	.128	—			
13 Cognitive/inattention ^a	-.581**	.116	-.512**	-.283	-.278	-.440*	-.455*	-.099	-.087	.094	-.14	.215	—		
14 Hyperactivity ^a	-.026	.031	-.420*	-.390*	-.379*	-.459*	-.503**	-.196	-.234	-.94	-.229	.751**	.434*	—	
15 ADHD ^a	-.228	.162	-.368*	-.270	-.297	-.391	-.424*	-.189	-.258	-.95	-.312	.366	.765**	.731**	—

^aTeacher Rating Scale.

* $p < .05$; ** $p < .01$.

when backward digit recall scores were partialled out, digit span scores remained significantly associated with action sequence recall, $r = .471$, $p < .05$ and action span, $r = .500$, $p < .01$, but not with either action feature recall, $r = .261$ or item recall, $r = .318$ ($p > .05$ in both cases). On balance, therefore, the strongest and most specific links were found between backward digit recall and performance in the action recall condition.

On the basis of our previous findings, it was anticipated that the working memory measure—backward digit recall—would also be significantly associated with scores on the cognitive problems/inattention subscale of the Teacher Rating Scale. This prediction was strongly upheld, $r = -.512$, $p < .01$, with 27 df. Backward digit recall scores were also significantly correlated, although less highly, with the other three subscale scores: oppositional, $r = -.373$, $p < .05$; hyperactivity, $r = -.420$, $p < .05$ and ADHD index, $r = -.368$, $p < .05$. However, given the small proportions of children obtaining atypically high scores (in excess of 55) on the oppositional and hyperactivity subscales as noted above, it is advisable not to draw strong conclusions concerning links between working memory ability and atypical behaviour in these domains on the basis of the data from these two measures.

Discussion

Working memory ability was highly associated with the accuracy with which children could perform verbal instructions involving sequences of actions. This result is consistent with our previous observations that children with low working memory scores struggle to follow instructions in a classroom context, and also with findings reported by Engle et al. (1991) that children's working memory scores were closely associated their abilities to follow complex task instruction that mimicked regular classroom activities. The instructions paradigm employed in the present study involved the manipulation of concrete three-dimensional objects that were highly familiar to the children, and used simple language to describe the required actions. Task difficulty was manipulated only by extending the number of steps in the instruction sequence to the point at which the child could no longer perform accurately. The close link between working memory and the ability to follow instructions demonstrated here is therefore unlikely to reflect the mediating impact of working memory on language comprehension. The stronger association between performance in the action condition and backward than forward digit recall indicates that the crucial constraint is not simply passive storage of the instruction, but rather the formation, maintenance and accessibility of the representation of the required action sequence in working memory during the course of the performed action. This interpretation is supported too by the absence of a corresponding association between backward digit recall and mere repetition of the instruction sequence.

As in Study 1, working memory ability was highly associated with atypically elevated levels of problem behaviours associated with the cognitive problems/inattention subscale of the Teacher Rating Scale. In this sample, significant although weaker links were also found between working memory and the other three subscale—oppositional, hyperactive and ADHD index scores—although it is notable that for the former two measures, it fell well within the typical range for almost all of the children.

GENERAL DISCUSSION

Two studies investigated the extent to which the working memory skills of 5- and 6-year-old children were associated with performance on laboratory analogues of

classroom activities on which children with low working memory scores had previously been observed to frequently fail (Gathercole et al., 2006). In Study 1, children with low and typical working memory scores performed two tasks that placed significant demands on both processing and storage. One task involved detecting and recalling rhyming words in spoken poems, and the other involved listening to a sentence and counting the number of words. In both cases, the low memory children performed more poorly than those in the average memory group on recall—of the rhyming words on one task, and of the sentences in the other. Importantly, their recall decrements did not appear to be a consequence either of problems with the processing element of the task—their abilities to detect rhymes, and to count the words—or of differences in general cognitive abilities as indexed by non-verbal reasoning performance. Instead, in both cases, the factor that most strongly predicted task performance was working memory ability.

According to one influential theoretical perspective on working memory, the processing and storage needs of an activity compete for a single limited working memory resource, leading to a trade-off between processing efficiency and storage (e.g. Case, Kurland, & Goldberg, 1982; Daneman & Carpenter, 1980). By this account, the links found between the complex memory and the recall measures from the two classroom analogue tasks employed in Study 1 could be a consequence of reduced processing efficiency of the children with low memory scores, which may diminish the resources available for the storage element of the tasks. The data provide little direct support for this view. Individual variation in the relevant measures of processing efficiency did not eliminate the recall differences found between the low and average working memory groups, consistent with other recent investigations of the relationship between processing and storage in complex memory span tasks (Bayliss et al., 2003; Saito & Miyake, 2004). The present findings indicate that the constraints on performance of the children with lower working memory scores relate to contributions of the central executive to activities other than processing, such as the coordination of processing and storage to meet task demands. Note also that the persistence of group differences in performance on the sentence counting and rhyme identification tasks in Study 1 when differences associated with scores on Raven's Coloured Matrices Test (1977) were controlled indicates that working memory performance cannot simply be equated with fluid intelligence, as argued by some theorists (e.g. Engle, Kane, & Tuholski, 1999).

Study 2 established that the ability to perform simple sequences of actions following verbal instructions was also strongly associated with working memory skills. The accuracy with which children followed instructions to manipulate sequences of dual-attribute (colour and form) objects was highly associated with a measure of working memory ability, backward digit recall. This feature of the results reinforces observational reports that children with poor working memory scores had difficulty in following lengthy instructions in the classroom (Gathercole et al., 2006; see also, Engle et al., 1991). Action performance was also associated with the forward digit recall measure of verbal storage, although in this case the link was not significant for the more sensitive item- and feature-based measures of recall that credited partially correct responses once the other measure (backward digit recall) was statistically controlled. Neither of the memory measures were significantly associated with the accuracy with which the children repeated the instruction, suggesting that performing a verbal instruction of this kind imposes the additional demands on working memory above and beyond its immediate recall.

The finding that the closest links were obtained between children's abilities to follow lengthy instructions and their backward digit recall scores indicates that performance on

this task is not simply constrained by passive verbal storage capacity. According to the working memory model introduced by Baddeley and Hitch (1974), forward digit recall is supported by the phonological loop, a component of working memory specialised for the storage of phonological material and is subject to rapid decay. Performance on complex memory tasks such as backward digit recall that require manipulation of the memory items in addition to their storage appears to depend both on the phonological loop and the central executive, a limited capacity attentional component that contributes to processing (Alloway et al., 2006; Bayliss et al., 2003; Kane, Hambrick, Tuholski, Wilhelm, Payne, & Engle, 2004). By this account, the present results indicate that following lengthy verbal instructions taps both the phonological loop and the central executive sub-components of working memory. It seems likely that a key element of the activity that requires support from the central executive is the need of the child to keep track of their place in the instruction sequence over the course of its performance.

One unexpected finding from Study 2 was that children were more than twice as accurate in performing than repeating the instructions. Because the repetition condition allowed the child to recall the instruction immediately rather than have to maintain its representation over the longer period during which the action was performed, it had been anticipated that superior performance would be found in the this condition. The converse advantage to action over repetition of the instruction has some correspondence with the widely documented subject-performed task (SPT) effect, in which recall of memory sequences is superior if the participant has actually performed an action than if they have read or heard the corresponding instruction, or even seen the experimenter perform it (see Nyberg, Persson, & Nilsson, 2002, for review). A variety of theories of the SPT effect have been advanced, including proposed benefits of physical actions for generating a distinctive motor programme (Engelkamp, 1998), obligatory semantic coding (Zimmer & Engelkamp, 1999) and enhanced integration of action-object combinations into a unitised whole (Mangels & Heinberg, 2006; Rumiati & Tessari, 2002). The present action advantage is distinguished from the SPT effect in that it manipulates the modality of recall rather than encoding; the verbal input form of the instructions is identical in both cases. Because theoretical accounts of the SPT effect typically focus on beneficial consequences of action at encoding, they do not strongly predict the present action advantage at recall. However, one possibility is that because the trials in the action and repetition conditions were blocked, participants may have constructed qualitatively different representations to guide performance in the two conditions. In particular, instructions in the action condition may have been encoded in the form of motoric or spatial representations that link the objects and actions, whereas a verbal representation may have been sufficient for immediate repetition of the sentence. The central executive is widely believed to be capable of handling information in any modality, and may play a crucial role in the construction of such relatively complex codes, superior recall of which may be due to a variety of factors such as greater distinctiveness or more effective preservation of sequential links between the actions. This account is at present speculative in nature, with further research needed to clarify the origin of what appears to be a substantial and previously unreported effect on immediate memory performance.

A secondary aim of this research was to investigate whether children with low working memory showed evidence of elevated levels of inattentive behaviour, as might be predicted from close links that have recently been reported between working memory deficits and the inattentive dimension of ADHD (Aronen et al., 2005; Klingberg et al., 2005; Martinussen & Tannock, 2006). Consistent with this hypothesis, children with low working memory

scores in both studies were rated by teachers as having particularly high levels of problem behaviours on the cognitive problems/inattention subscale of the Conners' Teacher Rating Scale (2001). One interpretational problem is that different items on this subscale relate to two distinct aspects of problem behaviours: poor academic achievement in the areas of reading, writing and arithmetic and inattention associated with short attention spans and high levels of distractibility. This raises the possibility that the high scores of the low working memory children simply reflect the well-documented findings that these individuals typically make poor academic progress (e.g. Gathercole et al., 2004; Swanson & Beebe-Frankenberger, 2004). Item analysis of the Teacher Rating Scale data for a separate sample of children with low working memory scores conducted by Gathercole et al. (in press) suggest that this is not the case, and that these individuals are typically rated as frequently exhibiting classroom problems relating both to poor learning and to the symptoms of inattention. The present study therefore adds to the accumulating evidence linking working memory problems with inattentive behaviour at the descriptive level at least, both in children with clinical diagnoses of attentional disorders and in non-clinical samples (see also, Aronen et al., 2005). The strong association between the two areas of difficulty raises important issues about whether or not a common deficit underlies failures of both working memory and sustained focussed attention.

In summary, working memory ability was closely related to children's performance on three laboratory analogues of classroom activities on which children with low working memory scores had previously been observed to struggle. Although the three activities have few surface features in common, they all require the construction and maintenance of a representation of an input form (a poem, or a meaningful sentence or a sequence of instructions) that must be maintained while the child is engaged in effortful cognitive processing (detecting rhyming word pairs, or counting words or performing the actions in sequence). As these are the two defining features of working memory, the findings fit well with the working memory overload hypothesis (Gathercole et al., 2006), according to which the generalised learning difficulties commonly associated with poor working memory capacity arise from inability to meet the working memory demands of classroom learning activities. The resulting working memory overload is likely to result in task failures that will inevitably impair their rates of learning.

The interpretation has direct implications for classroom practice (Gathercole & Alloway, 2008). Reducing working memory loads in classroom tasks may provide an effective means of preventing task failures and the associated lost learning opportunities, and hence of improving the academic progress of these children. Ways of doing this could include ensuring that the children work with short rather than long sentences or sequences that material is used that is meaningful rather than arbitrary where possible in order improve long-term memory support, and that the child is taught to use external memory aids to support their weak working memory capacities.

REFERENCES

- Alloway, T. P. (2007) *Automated working memory assessment*. London: Harcourt 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.
- Aronen, E. T., Vuontela, V., Steenari, M.-R., Salmi, J., & Carlson, S. (2005). Working memory, psychiatric symptoms, and academic performance at school. *Neurobiology of Learning and Memory*, 83, 33–42.

- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–90). New York: Academic Press.
- Bartha, L., & Benke, T. (2003). Acute conduction aphasia: An analysis of 20 cases. *Brain and Language*, *85*, 93–108.
- Bayliss, D. M., Jarrold, C., Gunn, M. D., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General*, *132*, 71–92.
- Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*, 386–404.
- Conners, C. K. (2001). *Conners Rating Scale-Revised*. North Tunawanda, NY: Multihealth Systems.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.
- Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *9*, 561–584.
- De Renzi, E., & Faglioni, P. (1978). Normative data and screening power of a shortened version of the Token Test. *Cortex*, *14*, 41–49.
- Engelkamp, J. (1998). *Memory for actions*. East Sussex, UK: Psychology Press.
- 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.
- Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). New York: Cambridge University Press.
- Gathercole, S. E., & Alloway, T. P. (2008) *Working memory and learning: A guide for teachers*. London: Sage Publishing.
- Gathercole, S. E., Alloway, T. P., Kirkwood, H. J., Elliott, J., Holmes, J., & Hilton, K. A. (in press). Attentional and executive function behaviours in children with poor working memory. *Learning and Individual Differences*.
- Gathercole, S. E., Alloway, T. P., Willis, C. S., & Adams, A. M. (2006). Working memory in children with reading disabilities. *Journal of Experimental Child Psychology* *93*, 265–281.
- Gathercole, S. E., Lamont, E., & Alloway, T. P. (2006). Working memory in the classroom. In: S. Pickering, & G. Phye (Eds.), *Working memory and education* (pp. 219–240). US: Academic Press.
- Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from National Curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, *18*, 1–16.
- Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, M. C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology*, *88*, 121–151.
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working-memory capacity: A latent-variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*, 189–217.
- Klingberg, T., Frenell, E., Olesen, P. J., Johnson, M., Gustafsoon, P., Dahlstrom, 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.
- Mangels, J. A., & Heinberg, A. (2006). Improved episodic integration through enactment: Implications for aging. *Journal of General Psychology*, *133*, 37–65.
- Martinussen, R., & Tannock, R. (2006). Working memory impairments in children with ADD with and without comorbid language learning disorders. *Journal of Clinical and Experimental Neuropsychology*, *28*, 1073–1094.
- Muter, V., Hulme, C., & Snowling, M. (1997). *Phonological abilities test*. London: Psychological Corporation.
- Nyberg, L., Persson, J., & Nilsson, L.-G. (2002). Individual differences in memory enhancement by encoding enactment: Relationships to adult age and biological factors. *Neuroscience and Biobehavioral Reviews*, *26*, 835–839.

- Raven, J. (1997). *Progressive Matrices Test*. Oxford: Oxford Psycholinguistics Press.
- Rumiati, R. I., & Tessari, A. (2002). Imitation of novel and well-known actions—The role of short-term memory. *Experimental Brain Research*, 142, 425–433.
- Saito, S., & Miyake, A. (2004). On the nature of forgetting and the processing-storage relationship in reading span performance. *Journal of Memory and Language*, 50, 425–443.
- Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development*, 60, 973–980.
- Swanson, H. L., Ashbaker, M. H., & Lee, C. (1996). Learning disabled readers working memory as a function of processing demands. *Journal of Experimental Child Psychology*, 61, 242–275.
- Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology*, 96, 471–491.
- Swanson, H. L., & Saez, L. (2003). Memory difficulties in children and adults with learning disabilities. In H. L. Swanson, S. Graham, & K. R. Harris (Eds.), *Handbook of learning disabilities* (pp. 182–198). New York: Guilford.
- Zimmer, H. D., & Engelkamp, J. (1999). Levels of processing effects in subject-performed tasks. *Memory and Cognition*, 27, 907–914.

APPENDIX 1

Stimuli from the sentence counting and recall task, Study 1

Practice trials	
Sport is fun	Rabbits eat grass
Pigs have four leg	The girl has pink shoes
Test trials	
Birds fly	Mice like cheese
Children play in the park	Rainbows are colourful
Dad cuts the grass	Ducks swim
Pigs like to roll around in the mud	Sam likes to read
Beth likes to paint pretty pictures	Houses are made of bricks
My aunt has a big dog	Peter likes cake
My sister plays music on her piano	Dogs bark
Pets must be looked after	Santa comes at Christmas
When it rains I use my umbrella	I am scared of the dark
I like to play with my brother and sister	My cat likes to sleep near the fire

APPENDIX 2

Stimuli from the detecting rhymes in poem task, Study 1

Practice trials	
Birds sing, phones ring	I like cakes, my mum bakes
Test trials	
Blue fish, green dish	Cats scratch, rats bite
A cowboy hat, a baseball bat	My dog swims, my frog hops
My shoes are clean, I ate a bean	My cat is blue, and so are you
My friend threw a bun, I thought it was fun	
