

# Working memory deficits in children with low achievements in the national curriculum at 7 years of age

Susan E. Gathercole\* and Susan J. Pickering

*University of Bristol*

**Background.** Close links between children's capacities to store and manipulate information over brief periods have been found with achievements on standardised measures of vocabulary, language comprehension, reading, and mathematics.

**Aim.** The study aimed to investigate whether working memory abilities are also associated with attainment levels in the national curriculum assessments at 7 years of age.

**Sample.** Eighty-three children aged 6 and 7 years attending local education authority schools participated in the study.

**Methods.** Working memory skills were assessed by a test battery designed to tap individual components of Baddeley and Hitch's (1974) working memory model. Children were assigned to normal and low achievement groups on the basis of their performance on national curriculum tasks and tests in the areas of English and mathematics.

**Results.** Children with low levels of curriculum attainment showed marked impairments on measures of central executive function and of visuo-spatial memory in particular. A single cut-off score derived from the test battery successfully identified the majority of the children failing to reach nationally expected levels of attainment.

**Conclusions.** Complex working memory skills are closely linked with children's academic progress within the early years of school. The assessment of working memory skills may offer a valuable method for screening children likely to be at risk of poor scholastic progress.

Since the early 1990s, all state schools in the UK have been required to follow a national curriculum which is taught and formally assessed at three points in time. At 7, 11, and 14 years of age, children's achievements within key areas of the curriculum are

\*Requests for reprints should be addressed to Professor Susan E. Gathercole, Department of Experimental Psychology, University of Bristol, 8 Woodland Road, Bristol BS6 6YH, UK (e-mail: sue.gathercole@bristol.ac.uk).

classified with reference to nationally expected standards. The classification of attainment levels at this point has extremely important consequences for both the children and their schools: it provides the basis both for identifying individuals who are failing to make normal progress, and for evaluation of the quality of education delivered by schools.

Attainment levels on the national curriculum assessment are likely to have a variety of influences including the children's basic intellectual capacities, the education they receive, and social and emotional factors. As yet, however, the contributions of these different factors to national curriculum performance have not been the subject of systematic study. The aim of the present study was to provide an initial investigation of the possible contribution to national curriculum achievements of one important part of the cognitive system which is already known to have close links with the acquisition of knowledge and skill, working memory.

The term 'working memory' is widely used to refer to a mental workplace in which information can be temporarily stored and manipulated in order to support ongoing complex cognitive activities (e.g., Anderson, 1983; Baddeley & Hitch, 1974; Daneman & Carpenter, 1980; Just & Carpenter, 1992). The specific model of working memory that guided the current project was advanced initially by Baddeley and Hitch (1974), and has been further elaborated on the basis of convergent evidence from experimental, developmental, neuropsychological and neuroanatomical studies over the past 25 years (see Gathercole & Pickering, *in press*, for review). The model is composed of three main components: the central executive, and two slave systems known as the phonological loop and the visuo-spatial sketchpad. The central executive is believed to support a variety of activities including controlling the flow of information through working memory, the retrieval of knowledge from long-term memory, the control of action, and the scheduling of multiple concurrent cognitive activities (e.g., Baddeley, 1986, 1996; Baddeley & Hitch, 1974; Baddeley, Emslie, Kolodny, & Duncan, 1998). Other models of complex working memory have characterised it as a system fuelled by a limited capacity resource which can be flexibly deployed to support ongoing processing and storage needs (e.g., Daneman & Carpenter, 1980; Just & Carpenter, 1992), and as activated portions of long-term memory controlled by attention (e.g., Cantor & Engle, 1992). Regardless of the detailed nature of the theoretical model, the scope and influence of the central executive (or complex working memory) is believed to extend to a wide range of complex cognitive activities (see Engle, Kane, & Tuholski, 1999, for review).

In contrast, the two slave systems of working memory operate within highly specialised processing domains. The phonological loop consists of a short-term store which retains verbal material in terms of its phonological characteristics, and which is subject to rapid decay. Decay of representations in the store can, however, be offset by a serial subvocal rehearsal process which may be articulatory in nature (Vallar & Baddeley, 1984). Developmentally, the phonological store appears to be in place by about 3 years of age, but the strategic rehearsal process does not emerge typically until about 7 years of age (Ford & Silber, 1994; Gathercole & Hitch, 1993).

The visuo-spatial sketchpad also appears to be composed of two separable subcomponents, one of which is visual, and one which is spatial (Hanley, Young, & Pearson, 1991; Farah, Hammond, Levine, & Calvanio, 1988; Logie, 1995; Logie &

Pearson, 1997; Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999). To a greater extent than the phonological loop, support for visuo-spatial short-term memory is dependent upon the central executive (Gathercole & Pickering, in press; Phillips & Christie, 1977; Wilson, Scott, & Power, 1987).

There is already a substantial body of evidence that working memory skills constrain the acquisition of a range of important complex abilities in childhood which are likely to have direct impact upon a child's success within school. Scores on measures of central executive function are closely associated with language and reading abilities (e.g., Siegel, 1994; Swanson, 1994; Yuill, Oakhill, & Parkin, 1989), arithmetic skills (Bull, Johnson, & Roy, 1999; Siegel & Linder, 1984; Siegel & Ryan, 1989), vocabulary acquisition (Daneman & Green, 1986), college entrance scores (Daneman & Carpenter, 1980; Jurden, 1995), and occupational success (Kyllonen & Christal, 1990). The phonological loop shares a more specific link with the acquisition of language, and in particular with the long-term learning of the sound patterns of new words (Baddeley, Gathercole, & Papagno, 1998; Cheung, 1996; Gathercole & Baddeley, 1989; Gathercole, Willis, Emslie, & Baddeley, 1992; Papagno, Valentine, & Baddeley, 1991; Papagno & Vallar, 1995; Service, 1992; Service & Kohonen, 1995). Marked deficits of phonological loop function in childhood are associated with specific language impairment, a severe and relatively common developmental disorder of language (Bishop, North, & Donlan, 1996; Gathercole & Baddeley, 1990; Montgomery, 1995). The visuo-spatial sketchpad, on the other hand, has been implicated in learning spatial routes and faces (Hanley *et al.*, 1991), and may also play an important role in the acquisition of arithmetic skills (Dark & Benbow, 1990).

The present study investigated whether the specific associations already established between aspects of working memory skill and particular domains of learning and achievement also extend to the more broad-ranging assessments of children's educational progress in the national curriculum. Working memory skills were assessed using a newly-developed working memory test battery designed to tap functioning in the central executive, phonological loop, and visuo-spatial sketchpad (Pickering & Gathercole, 1999). The test battery was administered to a large cohort of 6- and 7-year-old children (Gathercole & Pickering, 1999), which included the participants of the present study. Interrelationships between test scores established high internal validity for the central executive and phonological loop measures in particular. The test battery also has good external validity: in line with previous findings, the phonological loop measures shared highly specific links with vocabulary knowledge, whereas the central executive tasks showed more pervasive associations with achievements in the areas of language, literacy, and mathematics.

All children participating in the study had either recently completed or were about to complete the first stage of national curriculum assessments, which take place towards the end of the school year in which children are either 6 or 7 years of age. The standardised task and test scores were used to identify children performing below the nationally expected levels for their age in English and mathematics. If working memory skills do indeed limit children's capacities to acquire knowledge and skills in the educational domain, children with low curriculum achievements would be expected to perform poorly on the measures of working memory function. Findings of specific links between subcomponents of working memory and low achievements in particular areas

of curriculum would provide valuable insights into the theoretical underpinnings of these relationships.

## Method

### *Participants*

Eighty-three children (32 boys and 51 girls) attending three state primary schools in Bristol, South-West England participated in the study. The working memory assessments of the children were conducted in either the summer term of Year 2 or the autumn term of Year 3 (1997); their mean age was 7 years 4 months ( $SD = 3.72$  months, range 6 years 9 months to 8 years 0 months). All children were tested individually in a quiet area or room within the school, in four sessions separated by at least one day. The order in which the tests were administered was held constant where possible. National curriculum measures were administered in the second half of Year 2 (see below for further details).

### *Working memory assessments*

The working memory battery is composed of 13 tests designed to tap the three subcomponents of working memory (Pickering & Gathercole, 1999). Gathercole and Pickering (in press) supply examples of the stimuli employed in all unpublished tests. Details of their administration are provided below.

### *Phonological loop*

In the *forward digit recall* test, the experimenter spoke aloud sequences of digits which the child attempted to recall in the same sequence. Lists were constructed by sampling randomly and without replacement from the digits ranging from 1 to 9, and were presented at the rate of one digit per second. A typical digit sequence is 5, 1, 3. A maximum of four lists were presented at each length, with list length increasing by one if a child correctly recalled three lists at a particular sequence length. Testing commenced with two-digit lists, and continued until two or more lists of a particular length were recalled incorrectly. The maximum list length at which at least three lists were correctly recalled was calculated for each child. A test-retest reliability correlation coefficient for digit span of .68 was obtained in a study of 70 4- and 5-year-old children (Gathercole, 1995).

The *Children's Test of Nonword Repetition* (Gathercole & Baddeley, 1996) consists of 40 nonwords (10 each containing two, three, four, and five syllables) which are auditorily presented on an audio cassette recorder. Examples of nonwords include *woogalamic* and *skiticult*. The child was asked to repeat each nonword, and each repetition response was scored as either correct or incorrect (in the case of any phonological deviation from the target form). The maximum total score on this test is 40. Split-half reliability for this test is .66 (Gathercole & Baddeley, 1996), and test-retest reliability is .77 (Gathercole, 1995).

The *serial recognition of words and nonwords* was also tested. In each test, pairs of lists containing the same items in either identical sequence or with an adjacent transposition for a pair of the items were spoken by the experimenter (Gathercole, Service, Hitch, Adams, & Martin, 1999). The child's task was to judge whether the two lists were the

same or different. The items were spoken at a rate of one item per 750 milliseconds, with an extra 750 ms delay interpolated between the last item of the first list and the first item of the second list. Four sets of lists were presented at each list length of three, four, and five items, and in each set two list pairs are different and two pairs were the same. All lists were composed of one syllable stimuli (words or nonwords according to the version of the task) that were constructed from a restricted pool of 22 phonemes (11 consonants and 11 vowels). Only common words considered likely to be familiar children were included in the words lists.

At the beginning of the task, the child listened to a practice pair of lists (e.g., *cat, house, lid . . . cat, house, lid*), and was asked if the two lists were the same as one another, or different. In the next practice trial the same initial list was presented, but the second list transposed two of the items. Again, the child was asked to judge whether the pair was same or different. When the experimenter was satisfied that the child understood the task, the test trials commenced with four trials of lists pairs containing three items. The child then received four lists containing four items, and four lists containing five items. The number of correctly identified list pairs at each list length was scored. The maximum total score was 12; chance level of responding was .5.

The final two phonological loop tests involved *serial recall of words and nonwords*. On each trial, the child was asked to recall a sequence of items (e.g., *chin, led, bag*) spoken by the experimenter. Each word in the list was spoken at the rate of one item per 750 milliseconds. Following a practice trial, four lists were presented at each of the list lengths of three, four, and five items. However, as recall accuracy for the five-item lists was found to be extremely poor, only the number of three- and four-item lists correctly recalled (in terms of both items and order) were scored for each child. Thus, a maximum score of 8 was possible from the two list lengths retained for analysis.

### *Visuo-spatial memory*

All four visuo-spatial subtests were administered using a span procedure in which difficulty level increased across blocks of trials up to the point at which two or more trials are incorrectly recalled. A maximum of four trials was presented at each difficulty level in each case.

The *static and dynamic matrices* tests were both presented using an Apple Macintosh Powerbook 5300cs. The child was initially familiarised with the task requirements using practice items composed of two-dimensional matrices of squares, presented on the computer screen. On the first test trial of the static matrices test, a 2 by 2 matrix appeared in the centre of the computer screen containing two filled squares and was displayed for 2 seconds. This matrix was then replaced on the screen by a 2 by 2 matrix containing no filled squares, and the child was asked to identify the location of the filled squares in the original matrix by pointing at the screen. The examiner clicked each block in the matrix selected by the child in order to record the response. In the dynamic matrices version of the test, squares changed from white to black on for .5 seconds in sequence, with a .5 second delay between items. The children were asked to recall both the location and the order of the sequence of black squares.

If the child correctly recalled the location (or order and location) of the target items in at least three trials at this initial matrix size, the matrix increased to a 2 by 3 grid containing three filled squares. Testing continued until two or more errors were made in

one block of four trials, with two extra squares being added at each new difficulty level. On each test, the number of trials at which the child correctly recalled all target items (in correct sequence, in the case of the dynamic matrices test) was calculated. Test-retest reliability coefficient of .79 was obtained for the static matrices test, and .40 for the dynamic matrices subtests.<sup>1</sup> The low level of reliability for the dynamic measure indicates that further test refinement is needed to improve the psychometric properties of the test.

In the *static and dynamic mazes* tests, the child viewed a series of two-dimensional mazes which increased in complexity across trials. In the static mazes test, the child was given a response booklet containing pages of empty mazes. The first four mazes are composed of a stick person surrounded by two walls, each of which has two entry points. The child was instructed to look at a maze displayed by the experimenter for 2 seconds, on which is marked a red line extending from the outside of the maze to the central figure. The child was then asked to recall the route by drawing it on the corresponding maze in the response booklet. Following three or more correct responses in the four mazes at this difficulty level, the child progressed to mazes containing three walls, and so on.

In the dynamic mazes test, the route to be recalled was traced by the examiner's finger on the maze in full view of the child. Again, the task was to recall the route by drawing it into the response booklet. In both mazes tests, testing ceased when the child failed to recall two or more mazes at a particular difficulty level, and the total number of correct trials was scored for each child. Test-retest reliability coefficients for the static and dynamic subtests are .53 and .56, respectively.

#### *Central executive*

Three central executive tests were administered, each employing a span procedure with four trials at each difficulty level. The discontinuation criterion was failure to recall two or more items at a particular difficulty level. The *listening recall* subtest was based on the listening span procedure originally developed by Daneman and Carpenter (1980). On each trial, the child listened to a series of sentences, judged the veracity of each sentence in turn, and then recalled the final word of each of the sentences in sequence. This modified version of the listening span task was constructed using sentences from the Silly Sentences task (Baddeley, Emslie, & Nimmo-Smith, 1992) such as 'Frogs have long ears', and additional sentences of a similar kind which were constructed for the purposes of the present study.

Following two practice trials, each child was given four trials each consisting of two sentences. An example of a typical pair of lists was *Oranges live in water* and *Pigs have curly tails*. The number of sentences in each trial increased by one every four trials until the child incorrectly recalled two or more trials at a particular list length. The total number of correct trials (in which the child recalls both the words and their correct sequence) was scored. The test-retest reliability coefficient for the listening span task was .62.

The *counting recall* test was based on a procedure originally developed by Case, Kurland, and Goldberg (1982). On each trial, the child was required firstly to count and say aloud the number of dots presented in a series of arrays (providing a check on the number which the child believed to be present), and then to recall the dot totals in the

order that the arrays were presented. Each array consisted of a rectangle containing either three, four, five or six red dots. A practice booklet was used to familiarise children with the task. Testing began with trials containing two arrays of dots, and the number of arrays was increased by one every four trials. The total number of trials in which the numbers of dots were recalled in the correct order was scored. Low test-retest reliability was found for this measure,  $r = .15$ .

The *backwards digit recall* test employed the same procedure as the digit recall test in all respects except that the child was required to recall the sequence of spoken digits in reverse order. Thus, the sequence 8, 5, 2 would be correctly recalled as 2, 5, 8. Two practice trials were given in order to ensure that the child understood the concept of 'reverse'. Test trials commenced with four trials containing two digits, followed by lengthier sequences if three or more lists were correctly recalled. The total number of lists correctly recalled was scored. Split-half reliability for the WISC-III UK version of digit span, which includes both forwards and backwards recall of digits, is .85 (Golombok & Rust, 1992).

#### *National curriculum assessments*

The schools provided Key Stage 1 attainment levels in English, mathematics, and science for all children. There were two kinds of assessment: teacher assessments, and tasks/tests. The teacher assessments, completed by early July of 1997, were designed to provide a rounded judgment of children's performance across a range of contexts, and were guided by detailed level descriptions supplied by the School Curriculum and Assessment Authority (SCAA)/Qualifications and Curriculum Authority (QCA). There were teacher assessments in each of the three curriculum areas which were classified at the following levels: W (working towards level 1), 1 (below nationally expected standard), 2 (national expected standard), 3 (above nationally expected standard).

The tasks and tests used materials and methods developed and validated by SCAA/QCA (1997), and were intended to provide a snapshot of attainment at the end of the key stage which complement the broader-based teacher assessments. Tasks (reading, writing, and mathematics) were administered between January and June of 1997; tests (reading comprehension, spelling, and mathematics) were given in May 1997. The spelling and reading comprehension tests were only administered to children reaching level 2 at least on the reading task, so there were no level W or level 1 classifications on either of the tests. Spelling performance was classified as either L (falling below level 2), level 2 or level 3. For the remaining tasks and tests, level 2 performance was subdivided into three grades (a, b, c) in order to provide a finer discrimination within this average ability range into which most children fell. Children with scores at the high end of level 2 were classified as 2a, those with average achievements within the level as 2b, and those performing at the lower end of the level range were classified as 2c. The administration and marking of tests and tasks were audited and verified by local education authorities.

The aim of the present study was to use the national curriculum assessments to identify children judged to be failing to achieve normal levels of curriculum attainment. We chose to base our identification of children with low levels of achievement on the task and test results rather than the teacher assessments because whereas the psychometric properties of the teacher assessments were unknown the tasks and tests were fully standardised. Based on pre-test samples in excess of 2,000 children conducted

by NFER (DfEE, 1997), the following Cronbach's  $\alpha$  values were obtained: spelling test, .94; reading comprehension tests, .87-.77; mathematics test, .88. These values indicate satisfactory levels of reliability in each case.

**Table 1.** Percentages of children achieving each level/grade on the tasks and tests at Key Stage 1 (national figures for 1997 shown in parentheses)

Area Test/task	Attainment category								
	x	L	W	1	2	2c	2b	2a	3
English:									
Reading task	—	—	6 (3)	18 (16)	—	7 (18)	28 (20)	20 (16)	20 (26)
Reading comprehension test	24 (19)	0 (5)	—	—	—	16 (17)	23 (17)	17 (14)	20 (26)
Writing task	—	—	8 (6)	12 (13)	—	25 (33)	35 (27)	14 (14)	5 (6)
Spelling test	18 (7)	13 (30)	—	—	55 (47)	—	—	—	13 (14)
Mathematics:									
Test	—	—	4(2)	22 (14)	—	20 (18)	30 (20)	18 (25)	6(20)

x Pupils who were not entered for the reading comprehension and/or the spelling test because they obtained below level 2 on the reading task

L Pupils who failed to reach level 2 on either the reading comprehension or spelling test.

The percentages of children achieving each attainment level/grade in the tasks and tests, with national data for comparison, are shown in Table 1. The distribution of children across levels corresponded fairly closely with the national data for writing and spelling, although proportionately more children in our sample performed at the lower attainment levels on the reading task, reading comprehension test, and mathematics test. Note that as a consequence of our reliance upon the task and test results, no classification of children in science was provided as this was based solely on teacher assessment. In fact, the science teacher assessments corresponded very closely with teacher assessments in English and mathematics: all of the 16 children who attained either level W or 1 in science also failed to reach level 2 in the two other areas of the curriculum. There is therefore little evidence that low achievement levels in science at this educational stage reflect specific learning difficulties.

An analysis of the validity of the attainment levels across the four English tasks and tests was conducted from our own dataset. To do this, attainment levels were assigned the following numerical values: W (0), L(1), 1 (1), 2c (2), 2b (3), 2 (3), 2a (4), 3 (5). Children who did not qualify for administration of the reading comprehension and spelling tests (classified x in Table 1) were scored as 0. The resulting mean scores (with standard deviations) were: reading task, 3.00 (1.55), reading comprehension test, 2.70 (1.81), writing task, 2.49 (1.26), spelling test, 2.37 (1.53). Reliability was assessed by computing correlation coefficients for each pair of measures. The 20 children failing to reach level 2 on the reading task were excluded from analyses of the comprehension and spelling scores because they did not qualify for administration of these tests. The following correlation coefficients were obtained: reading and writing,  $r(81) = .84$ ,  $p < .001$ ; reading and reading comprehension,  $r(61) = .85$ ,  $p < .001$ ; reading and spelling,  $r(61) = .57$ ,  $p < .001$ ; reading comprehension and spelling,  $r(61) = .52$ ,

$p < .001$ ; writing and comprehension,  $r(61) = .69$ ,  $p < .001$ ; writing and spelling,  $r(61) = .53$ ,  $p < .001$ .

These close links between scores, particularly for reading, reading comprehension, and writing, indicate that the assessments provide valid indicators of the children's competence in English. Spelling scores showed rather lower, although still highly significant, associations with other attainment levels, suggesting that it may reflect a relatively distinct ability within the domain of literacy.

The total numerical score for the reading, reading comprehension, and writing measures provided a single index of attainment in English for the purposes of this study. The spelling measure was not included in this composite score due both to the absence of fine discrimination between the level 2 category afforded by the other measures and to the lower correlations found between this and the other three measures.

Children obtaining mathematics attainment levels below 2 (W or 1) were identified as low achievers in this curriculum area for the purposes of our analyses. Two methods of selecting low achieving children in English were compared. The first method involved scoring the highest attainment level in the reading, reading comprehension, and writing measures for each child. Children whose highest level fell below level 2 were classified as low achievers in English. By giving credit for the child's strongest area of performance within the domain of English, this method is likely to minimise the inclusion of children in the low achievement category whose actual competence is either marginal or average. The second method used the total numerical scores for the reading, reading comprehension, and writing measures described above. Children with scores below 3 were classified as low achievers. Both methods succeeded in identifying the same children as belonging to the low and normal achievement groups.

In total, 23 children were identified as having low achievements on their Key Stage 1 assessments. Of these, 13 children were classified as low achievers in both English and mathematics (7 boys and 6 girls, mean age 7 years 1 month), 8 were identified as low achieving in mathematics only (3 boys and 5 girls, mean age 7 years 1 month) and the remaining 2 children were low achievers in English only (1 boy and 1 girl, mean age 6 years 11 months). The remaining 60 children did not fall into either low achievement category (21 boys and 39 girls, mean age 7 years 5 months). There were no significant differences in the ages of the children between any pairs of groups.

## Results

The distributions of the working memory variables were examined for their fit with the assumptions of multivariate analysis. All 13 memory variables were found to be reasonably normally distributed, with skewness and kurtosis values close to zero. As no multivariate outliers were found using Mahalanobis distance with  $p < .001$ , available data for all children were included in the analyses.

The degree to which the test scores correlated with age varied considerably. As reported by Gathercole and Pickering (in press), chronological age in months correlated significantly with the following measures: backwards digit recall,  $r(81) = .27$ ,  $p < .05$ ;

counting recall,  $r(81) = .33, p < .01$ ; dynamic matrices,  $r(81) = .23, p < .05$ ; static mazes,  $r(81) = .40, p < .01$ ; dynamic mazes,  $r(81) = .29, p < .01$ ; word matching,  $r(81) = .28, p < .05$ .

The performance of the normal and low achievement national curriculum achievement groups is summarised in Table 2. Effect sizes, based on comparisons of the group with no areas of low achievement with each low achievement group, are also shown in Table 2. This measure is the difference in the  $z$ -scores (calculated for the group as a whole) between the two groups, so that an effect size of 1.00 corresponds to a deficit of one standard deviation in the low achievement group compared with the normal achievement group. Differences between means which are significant by univariate  $F$ -tests at the 1% level (adopted due to the relatively large numbers of comparisons being made) are also indicated in the table.

Performance was generally lower for the low achievement groups than the children with no areas of low achievement, with the most marked differences found in the children with low achievements in English and/or mathematics, particularly for the central executive and visuo-spatial tasks. In terms of effect sizes, the deficits were even stronger for both the children with low achievements in mathematics only and the children with low achievements in both areas of the curriculum, but due to the relatively small sizes of these two groups the differences were generally non-significant. No comparisons of means were conducted for the low English achievement only children, due to small numbers. As a consequence of the small group sizes for all low achievement groups other than those with low achievements in either English and/or mathematics, subsequent analyses compared only the two largest contrasting groups: the children with no areas of low achievement ( $N = 60$ ) and children with one or more area of low achievement ( $N = 23$ ).

The working memory dataset includes a mix of both correlated and uncorrelated variables, with measures clustering principally around two factors: a central executive factor on which all of the central executive subtests load and three of the visuo-spatial memory subtests (excluding static matrices), and a phonological loop factor associated uniquely only with the phonological loop measures (see Gathercole & Pickering, in press, for further information). Although this profile indicates that the visuo-spatial measures included in the test battery do not tap a separate factor to the central executive, in order to aid theoretical interpretation of the data it was considered useful to conduct separate analyses for the subtests associated with each of the three components of the working memory model. Accordingly, separate MANOVAs were performed for the sets of measures associated in previous research with each component of working memory, as a function of achievement group. The three central executive tasks were entered as dependent variables in one MANOVA, the six phonological memory test scores in another MANOVA, and the four visuo-spatial tasks in a third analysis.

Significant achievement group effects were found in the MANOVAs performed on the central executive measures,  $F(3,79) = 8.97, p < .001$ , and on the visuo-spatial memory scores,  $F(4,77) = 8.09, p < .001$ , although not on the phonological loop tests,  $F(6,76) = 1.78, p > .05$ . Individual tests which yielded significant differences on the univariate  $F$ -tests were all three central executive tasks (backwards digit recall, counting recall and listening recall,  $p < .001$ , in each case), two visuo-spatial measures (dynamic

**Table 2.** Performance of the national curriculum achievement groups on the working memory tests

	Area/s of low achievement on national curriculum tests															
	None			English only			Maths only			English & maths			English &/or maths			
	Mean	SD	N = 60	Mean	SD	Effect size†	Mean	SD	Effect size	Mean	SD	Effect size	Mean	SD	Effect size	
<b>Working memory comp</b>																
<i>Central executive:</i>																
Backwards digit recall	7.50	2.43	5.00	1.41	1.02	0.46	6.37	1.06	0.89	4.85	1.99	1.08*	5.39	1.78	0.89*	
Listening recall	3.48	1.92	2.00	2.83	0.79	0.93	1.75	0.89	0.93	1.85	0.90	0.87	1.83	1.03	0.88*	
Counting recall	7.68	2.76	6.50	2.12	0.41	1.01	4.75	1.49	1.01	5.54	3.13	0.74	5.35	2.55	0.81*	
<i>Phonological loop:</i>																
Nonword repetition	25.67	5.97	23.00	14.14	0.42	0.89	20.00	5.92	0.89	23.62	6.60	0.32	22.30	6.86	0.53	
Word recall	4.45	1.83	5.00	1.41	-0.31	0.89	2.87	1.64	0.89	3.69	1.25	0.43	3.52	1.47	0.52	
Nonword recall	1.35	1.42	1.00	1.41	0.27	0.08	1.25	1.39	0.08	1.00	0.70	0.27	1.09	1.00	0.20	
Forward digit recall	4.10	0.66	4.50	0.71	-0.61	0.53	3.75	0.71	0.53	3.77	0.60	0.50	3.83	0.65	0.41	
Word matching	9.70	1.61	10.00	0.00	-0.17	0.45	8.88	1.89	0.45	8.08	2.25	.90*	8.52	2.06	0.65*	
Nonword matching	8.32	1.99	8.50	0.71	-0.09	0.16	8.00	1.07	0.16	6.92	2.43	0.69	7.43	2.00	0.44	
<i>Visuo-spatial memory:</i>																
Static matrices	17.19	3.58	19.50	2.12	-0.59	0.91	13.62	5.26	0.91	15.38	3.91	0.46	15.13	4.47	0.52	
Dynamic matrices	10.24	3.00	6.50	0.71	1.13	0.86	7.37	2.50	0.86	7.46	3.97	0.84	7.35	3.27	0.87*	
Static mazes	8.48	2.80	5.50	2.12	0.99	1.11*	5.12	2.23	1.11*	6.16	3.13	0.77	5.74	2.72	0.89*	
Dynamic mazes	9.05	2.82	9.00	2.83	0.02	0.81	6.75	1.91	0.81	7.62	2.90	0.51	7.43	2.56	0.57	

† No significance tests were performed due to small numbers

\*  $p < .01$ , by univariate  $F$ -test

matrices and static mazes,  $p < .001$ , in both cases), and a single phonological measure, word matching ( $p < .01$ ).

Thus, children performing below nationally expected levels in their national curriculum assessments at Key Stage 1 showed marked deficits in particular on measures of both the central executive and visuo-spatial short-term memory. Next, the extent to which scores on the individual working memory measures could be used to predict children with at least one area of low achievement was explored using the method of discriminant function analysis. This method computes a weighted linear function of input variables which optimises correct assignment to groups, and so provided a means of identifying which working memory scores were the best predictors of poor curriculum achievement. In this analysis,  $z$ -scores for all 13 working memory subtest scores were used to predict membership of the normal achievement and low achievement (in either one or two areas of the curriculum) groups. Five variables emerged as significant predictors of group membership in this analysis (in order of entry into the discriminant function), yielding a single function (standardised coefficients shown in parentheses): static mazes (.45), listening recall (.34), dynamic matrices (.43), backwards digit recall (.29), nonword repetition (.31), and static matrices (.23). The  $\chi^2$  value of this function was 38.59 with 6 degrees of freedom,  $p < .001$ . This function resulted in the successful classification of 82.9% of the children. Correct group assignment was achieved for 49 of the 59 (83.1%) normal achievement children, and for 19 of the 23 (82.6%) children in the low achievement group.

Scores on a subset of the working memory measures from the test battery can therefore be used to identify with reasonable accuracy those children who fail to achieve expected levels of attainment on one or more areas of the national curriculum at 7 years. It is notable that the discriminating variables that emerge from these analyses draw from all three components of the working memory model, and so point to general rather than specific support for the acquisition of knowledge and skills in the broad educational context.

Finally, the extent to which a small subset of working memory scores could generate a single 'working memory risk' score which could be used to identify children with low achievements in one or more areas of the curriculum was investigated. The choice of which measures to include in this computation was initially guided by the results of the discriminant function analyses reported above. The best single combination was obtained by summing the  $z$ -scores for three measures: backwards digit recall, nonword recall, and nonword recognition. In terms of the proportion of children from the entire sample who were correctly assigned to their achievement categories using this score, the optimal cut-off score was  $-.62$  (below which children were assigned to the low achievement category). A total of 14 of the 23 (61%) members of the low achievement group were correctly identified using this criterion, and 92% of the children with normal achievement levels (giving a correct assignment value for the sample as a whole of 89%). A lower cut-off score was, however, much more effective at identifying children belonging to the low achievement group. A total of 20 of the 23 children (87%) in this group were correctly classified by applying a cut-off score of  $-.08$ ; 19 of the 60 (68%) normal achievement group were also correctly assigned to their category. Although this lower score gives a lower overall rate of correct group assignment across the sample (73%) than the higher cut-off score due to the higher false positive rate

(normal achievement children incorrectly classified as low achievement), it has the advantage of a much higher hit rate for the poor performance group and therefore may be more useful in screening terms for identifying children with some risk of curricular under-achievement.

### Discussion

This study provides substantial evidence for close links between children's performance on national curriculum assessments at 7 years of age and their working memory skills. Children who failed to achieve expected levels of attainment for their age in one or more areas of curriculum assessment performed poorly on tests of working memory function. The deficits were particularly severe for the central executive measures, each of which required the child to process and store verbal material simultaneously, and also for the tests of visuo-spatial memory. The pattern was less consistent for the phonological loop measures. Although low achieving children were significantly impaired on a subset of these measures, the differences did not extend across the full range of tasks.

These findings lend further weight to previous evidence that the central executive in particular plays a crucial role in the acquisition of complex cognitive abilities and skills such as literacy, comprehension, and arithmetic (e.g., Bull *et al.*, 1999; Swanson, 1994; Yuill *et al.*, 1989). The consistency of the impairments shown by the children with low levels of attainment in their national curriculum assessments on all three central executive measures lends considerable generality to this association. The link between poor achievements in the curriculum and scores on the visuo-spatial measures was rather more unexpected, as there is little evidence in the research literature that the visuo-spatial sketchpad plays a key role in scholastic learning in general across these three broad areas of the curriculum, although more specific links have been found with the development of mathematical ability (Dark & Benbow, 1990). Our recent analysis of the internal validity of the working memory battery employed in the present study may, however, cast some useful light on this association. When the factor structure of subtest scores was examined, we found that both the central executive and visuo-spatial measures loaded on a single factor (Gathercole & Pickering, in press). It may therefore be the case that many visuo-spatial tasks may not call upon a specialised and distinct component of working memory designed to support processing and storage within that domain, but instead depend upon the general purpose and highly flexible resources of the central executive (see also, Phillips & Christie, 1977; Wilson *et al.*, 1987). Perhaps, then, the associations found between scores on the visuo-spatial subtests and national curriculum performance reflects the role of the central executive in supporting achievements in these curricular domains.

The cause of the substantial impact of central executive capacities on the acquisition of knowledge and complex skills during the school years seems likely to lie in the use of working memory as a mental workspace in which products of ongoing processes can be stored and integrated during complex and demanding activities (Just & Carpenter, 1992). Such activities which are common in the classroom and which form the basis for making progress within areas of the curriculum include listening to another speaker, decoding an unfamiliar word whilst maintaining the meaning of previously decoded

text, writing whilst formulating the next part of a text, and engaging in mental arithmetic. In each case, the child is required to process new information and to integrate it with knowledge that has either been retrieved from long-term memory or is held temporarily within working memory. From this perspective, it may therefore be not too surprising that a child with a very restricted capacity to engage in such mental activities will fail to make normal progress within school.

There is little evidence that phonological loop limitations constrain curriculum progress within the early school years, although children with low achievements in one or more of the areas of assessment were poor at recognising word sequences. Although there are undoubtedly some variations in sensitivity of the individual subtests to the phonological loop, it seems unlikely that the unevenness of the link with curricular achievement is simply a consequence of this. Our assessment of the internal validity of the working memory battery indicated extremely good internal coherence for the full set of phonological loop measures, with all six measures loading significantly on a single construct (Gathercole & Pickering, 1999). We did, however, find that the phonological loop and central executive factors shared a moderate and highly significant degree of association, and this may explain some of the present findings. The phonological loop subtests which did effectively discriminate the curricular achievement groups may possibly have placed significantly greater demands upon the central executive than the other phonological subtests, and this could lie at the root of their sensitivity to school achievement.

The failure to obtain strong and consistent associations between the phonological loop measures and children's achievements in school may seem surprising in the context of recent evidence for the role played by the loop in the acquisition of language, and in vocabulary learning in particular (see Baddeley *et al.*, 1998, for review). It should, however, be noted that the majority of the children in the low achievement groups had failed to make normal progress in both English and mathematics. Only ten children had low achievements in a single area only: eight children performed poorly in mathematics, and the remaining two had problems in English. As these group sizes are too small to merit statistical consideration as separate groups, they were treated as a single low achievement group. Our data therefore provide valuable information on the role of working memory (and the central executive in particular) in supporting curricular progress generally, but do not assess the contribution of individual subcomponents of working memory to learning in more specific domains. To investigate this, it will be necessary either to work with much larger samples of children, or to study working memory profiles in children with more specific learning disabilities such as specific language impairment (Bishop *et al.*, 1996; Dollaghan & Campbell, 1998; Gathercole & Baddeley, 1990).

The present findings that central executive deficits in particular characterise children with curricular achievements falling below national levels for age on national curriculum assessments may have important practical theoretical implications. Using a score derived from only three working memory subtests (backwards digit recall, nonword repetition, and nonword recognition), we were able to predict with a high level of accuracy the children who failed to reach normal attainment levels in one or more areas of the curriculum. Although the present study investigated contemporaneous rather than prospective relationships between working memory and curriculum

learning, one possibility is that working memory assessments may play a useful role in screening for children at risk of educational underachievement. Many of these tests can be used in their current form with children considerably younger than 7 years of age (e.g., Ford & Silber, 1994; Gathercole & Adams, 1993), and we have recently modified the procedures used in the more complex central executive tasks to make them appropriate for 4-year-old children. It would be extremely valuable to discover whether such working memory assessments at school entry could provide accurate prospective indicators of failure to reach normal levels of attainment within the national curriculum at later points in the educational process.

There may be other benefits to using working memory assessments prospectively, too. State schools in England are now under a statutory obligation to provide baseline assessments of children's abilities at school entry at 4 years of age, and typical schemes in current usage focus on the child's mastery of basic concepts in particular knowledge domains such as number and print. Whilst undoubtedly providing a useful insight into the child's current abilities, these measures may be strongly influenced by the quality and quantity of environmental support for learning and structured tuition that a child has already received. In contrast, the working memory tests employed in the present study use stimuli and methods which are likely to be equally unfamiliar to all children, and appear to be relatively free from cultural and environmental experience (Campbell, Dollaghan, Needleman, & Janosky, 1997). Working memory measures may therefore provide fluid and sensitive indicators of the child's ability to acquire knowledge and understanding in key aspects of the curriculum which act as a useful supplement to knowledge-based methods of baseline evaluation.

### Acknowledgments

This research was supported by a programme grant awarded by the Medical Research Council of Great Britain to Alan Baddeley and Susan Gathercole. The authors wish to thank Melanie Hall and Simon Lloyd for their assistance in data collection, and to the children and staff of Parson Street Primary School and St Michael's C. of E. Primary School in Bristol, for their time and assistance which they generously contributed to this study.

### NOTE

<sup>1</sup> Test-retest reliability was assessed in 27 of the children participating in this study on the following subtests: matrices static and dynamic, mazes static and dynamic, listening span, and counting span. Following administration of the full working memory battery at initial testing, these five subtests were administered again to these children three weeks later.

### References

- Anderson, J.R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Baddeley, A.D. (1986). *Working memory*. Oxford: Oxford University Press.
- Baddeley, A.D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology*, 49A, 5–28.
- Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp.47–90). New York: Academic Press.

- Baddeley, A.D., Emslie, H., & Nimmo-Smith, I. (1992). *The speed and capacity of language processing (SCOLP) test*. Flemspton, Bury St Edmonds. Suffolk: Thames Valley Test Company.
- Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*, 158–173.
- Baddeley, A. D., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of memory. *Quarterly Journal of Experimental Psychology*, *51A*, 819–852.
- Bishop, D.V.M., North, T., & Donlan, C. (1996). Nonword repetition as a phenotypic marker for inherited language impairment: Evidence from a twin study. *Journal of Child Psychology and Psychiatry*, *33*, 1–64.
- Bull, R., Johnson, R.S., & Roy, J.A. (1999). Exploring the roles of the visuo-spatial sketchpad and central executive in children's arithmetical skills: Views from cognition and developmental neuropsychology. *Developmental Neuropsychology*.
- Campbell, T. F., Dollaghan, C., Needleman, H. & Janosky, J. (1997). Reducing bias in language assessment: Processing-dependent measures. *Journal of Speech, Language, and Hearing Research*, *40*, 519–525.
- Cantor, J., & Engle, R.W. (1993) Working memory capacity as long-term activation: An individual differences approach. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *19*, 1101–1114.
- Case, R. D., Kurland, M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*, 386–404.
- Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology*, *32*, 867–873.
- Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behaviour*, *19*, 450–466.
- Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*, 1–18.
- Dark, V.J., & Benbow, C.P. (1990). Enhanced problem translation and short-term memory: Components of mathematical talent. *Journal of Educational Psychology*, *82*, 420–429.
- Della Sala, S., Gray, C., Baddeley, A. D., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unwelding visuo-spatial memory. *Neuropsychologia*, *37*, 1189–1199.
- Department for Education and Employment. (1997). *Results of the 1997 National Curriculum assessments of 7 year olds in England*. London: HMSO.
- Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, *41*, 1136–1146.
- 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* (pp.102–134). Cambridge: Cambridge University Press.
- Farah, M.J., Hammond, K.M., Levine, D.N., & Calvanio, R. (1988). Visual and spatial mental imagery: Dissociable systems of representation. *Cognitive Psychology*, *20*, 439–462.
- Ford, S., & Silber K.P. (1994). Working memory in children: A developmental approach to the phonological coding of pictorial material. *British Journal of Developmental Psychology*, *12*, 165–175.
- Gathercole, S.E. (1995). The assessment of phonological memory skills in preschool children. *British Journal of Educational Psychology*, *65*, 155–164.
- Gathercole, S.E., & Adams, A. (1993). Phonological working memory in very young children. *Developmental Psychology*, *29*, 770–778.
- Gathercole, S.E., & Baddeley, A.D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, *28*, 200–213.
- Gathercole, S.E., & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology*, *81*, 439–454.

- Gathercole, S. E., & Baddeley, A. D. (1996). *The Children's Test of Nonword Repetition*. San Diego, CA: The Psychological Corporation.
- Gathercole, S. E., & Hitch, G. J. (1993). Developmental changes in short-term memory: A revised working memory perspective. In A. Collins, S. E. Gathercole, M. A. Conway, & P. E. Morris (Eds.), *Theories of memory* (pp. 189–210). Hove, UK: Erlbaum.
- Gathercole, S. E., & Pickering, S. J. (in press). Assessment of working memory in six- and seven-year old children. *Journal of Educational Psychology*.
- Gathercole, S. E., Wills, C., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, 28, 887–898.
- Gathercole, S. E., Service, E., Hitch, G. J., Adams, A.-M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence of the nature of the relationship. *Applied Cognitive Psychology*, 13, 65–77.
- Golombok, S., & Rust, J. (1992). *Manual of the Wechsler Intelligence Scale for Children – Third Edition UK*. Kent: The Psychological Corporation.
- Hanley, J.R., Young, A.W., & Pearson, N.A. (1991). Impairment of the visuo-spatial sketch pad. *Quarterly Journal of Experimental Psychology*, 43A, 101–125.
- Jurden, F. H. (1995). Individual differences in working memory and complex cognition. *Journal of Educational Psychology*, 87, 93–102.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122–149.
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability (is little more than) working memory capacity. *Intelligence*, 14, 389–433.
- Logie, R.H. (1986). Visuo-spatial processing in working memory. *Quarterly Journal of Experimental Psychology*, 38A, 229–247.
- Logie, R.H. (1995). *Visuo-spatial working memory*. Hove, UK: Lawrence Erlbaum.
- Logie, R.H., & Pearson, D.G. (1997). The inner eye and the inner scribe of visuo-spatial working memory: Evidence from developmental fractionation. *European Journal of Cognitive Psychology*.
- Montgomery, J. (1995). Examination of phonological working memory in specifically language-impaired children. *Applied Psycholinguistics*, 16, 355–378.
- Papagno, C., Valentine, T., & Baddeley, A.D. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language*, 30, 331–347.
- Papagno, C., & Vallar, G. (1995). Short-term memory and vocabulary learning in polyglots. *Quarterly Journal of Experimental Psychology*, 48A, 98–107.
- Phillips, W. A., & Christie, D. F. M. (1977). Interference with visualization. *Quarterly Journal of Experimental Psychology*, 29, 637–650.
- Pickering, S. J., & Gathercole, S. E. (1999). *Working memory battery*. Available from the authors at Department of Experimental Psychology, University of Bristol, England.
- SCAA/ QCA. (1997). *Key Stage 1 Assessment Arrangements*. London: HMSO.
- Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology*, 45A, 21–50.
- Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign-language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics*, 16, 155–172.
- Siegel, L.S. (1994). Working memory and reading: A life-span perspective. *International Journal of Behavioural Development*, 17, 109–124.
- Siegel, L.S., & Linder, B.A. (1984). Short-term memory processes in children with reading and arithmetic disabilities. *Developmental Psychology*, 20, 200–207.
- 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. (1994). Short-term memory and working memory: Do both contribute to our understanding of academic achievement in children and adults with learning disabilities? *Journal of Learning Disabilities*, 27, 34–50.

- Vallar, G., & Baddeley, A.D. (1984). Fractionation of working memory: Neuropsychological evidence for a short-term store. *Journal of Verbal Learning and Verbal Behavior*, *23*, 151–161.
- Wilson, J.T.L., Scott, J.H., & Power, K.G. (1987). Developmental differences in the span of visual memory for pattern. *British Journal of Developmental Psychology*, *5*, 249–255.
- Yuill, N., Oakhill, J., & Parkin, A. (1989) Working memory, comprehension ability and the resolution of text anomaly. *British Journal of Psychology*, *80*, 351–361.

*Received 28 August 1999; revised version received 9 September 1999*