Methods of Timepoint 1 (2017)

UK Data

A longitudinal study of mathematical development in primary schools from Year 1 to Year 3

Funded by the ESRC, UK, grant number ES/N014677/1

Table of Contents

1. Participants

Participants were 314 children aged between five- and six-years (M_{age}= 74.86 months, *SD*=3.75), 168 males and 147 females attending Year one at primary school. From the sample tested, 17 (5%) children were classed as EAL or bilingual by the class teacher. Children were recruited from both suburban and rural primary schools in areas across North and West Yorkshire. The ethnicity descriptions and rural/urban descriptions are taken from surrounding area 2011 census data. The school Free School Meal Percentage (FSM), and number in school, data are taken from 2017 school census data.

The National Average FSM in the last six years (i.e. 2011 - 2017) was 24.9%. The deprivation index is taken from the most current English Indices of Deprivation (2015, see Appendix 1). A description of each school can be found in Table 1.

Table 1: Background information on participating schools

School Number	Number of Children	School Type
1	39 (2 KS1 classes: One Reception/Y1 mixed class and one Y1/Y2 mixed class, only Y1 children tested)	Co_Ed Community Primary 280 in school. 4% FSM anytime in last six years[1] 93% White English, Scottish, Welsh, North Irish or British in local area. In 10% least deprived neighbourhoods using English Indices of Deprivation (2015) (decile score 10)
2	19 1 Year 1 class	Co_Ed Community Primary 139 in school. 13% FSM anytime in last six years[1] 96.5% White English, Scottish, Welsh, North Irish or British in local area. In 10% least deprived neighbourhoods using English Indices of Deprivation (2015) (decile score 10)
3	29 1 Year 1 class	Co_Ed Community Primary 210 in school. 5% FSM anytime in last six years[1] 87% White English, Scottish, Welsh, North Irish or British in local area. In 10% least deprived neighbourhoods using English Indices of Deprivation (2015) (decile score 10)
4	28 1 Year 1 class	Co_Ed Academy 170 in school. 20% FSM anytime in last six years[1] 97% White English, Scottish, Welsh, North Irish or British in local area. In 10% most deprived neighbourhoods using English Indices of Deprivation (2015) (decile score 1).
5	8 Mixed EYFS/KS1 Class (Reception and Year 1, only Y1 children tested)	Church of England controlled Co_Ed Primary. 65 in school. 23% FSM anytime in last six years[1] 96% White English, Scottish, Welsh, North Irish or British in local area. In 40% most deprived neighbourhoods using English Indices of Deprivation (2015) (decile score 4)
6	20 One Year 1 class	Community Co_Ed Primary. 322 in school. 4% FSM anytime in last six years[1]

		96% White English, Scottish, Welsh, North Irish or British in local area. In 30% least deprived neighbourhoods in England, using English Indices of Deprivation (2015) (decile score 8).
7	58 From Three Key Stage 1 (age 5-7) classes: two Y1 classes and one mixed Y1/Y2 class (Y1 children tested only)	Academy Co_Ed Primary. 226 in school. 4% FSM anytime in last six years[1] 95% White English, Scottish, Welsh, North Irish or British in local area. In 10% least deprived neighbourhoods using English Indices of Deprivation (2015) (decile score 10)
8	32 Two Year 1 Classes	Private non-selective school age range 2-18 536 in school. 0% recorded FSM anytime in last six years[1] 95% White English, Scottish, Welsh, North Irish or British in local area. In 20% least deprived neighbourhoods in England, using English Indices of Deprivation (2015) (decile score 9).
9	20 One Year 1 class	Church of England Voluntary Controlled Co_Ed Primary School 163 in school. 9% recorded FSM anytime in last six years[1] 98% White English, Scottish, Welsh, North Irish or British in local area. In 10% least deprived neighbourhoods in England, using English Indices of Deprivation (2015) (decile score 10).
10	43 Three mixed KS1 Classes (Y1/Y2 mixed classes, Y1 children tested); five children kept in Reception also tested).	Academy Co_Ed Primary School 317 in school. 13% recorded FSM anytime in last six years[1] 89% White English, Scottish, Welsh, North Irish or British in local area In 30% least deprived neighbourhoods in England, using English Indices of Deprivation (2015) (decile score 8)
11	24 One Year 1 Class	Academy Co_Ed Primary School 225 in school. 27% recorded FSM anytime in last six years[1] 89% White English, Scottish, Welsh, North Irish or British in local area In 30% least deprived neighbourhoods in England, using English Indices of Deprivation (2015) (decile score 8).

[1] Data taken from 2017 school census (National Statistics, 2017)

Key Stage 1 (KS1) is the legal term for two years of schooling between Year 1 and Year 2 when pupils are aged 5 to 7 years. Early Years Foundation Stage (EYFS) refers to providers of care for children under the age of five, which is the age of compulsory education in the UK.

Testing took place between May and July 2017. The head teachers in all participating schools gave consent and letters were sent to parents of all children who were eligible to take part in the study informing them of the study. Parents were given the opportunity to opt their child(ren) out of being tested, with the exception of one school who decided on an opt-in approach. Children with developmental disorders and/or neurological disorders were not excluded from the study. 25 children were included on the schools SEN register. The study was approved by the University of York Psychology Department Ethics Committee (Reference number 559).

2. Materials and Stimuli

Materials used consisted of a number of standardised and non-standardised tasks which were used during group and individual testing sessions. Tasks measured a number of cognitive constructs and are detailed in Table 2, in the order in which they were completed.

Table 2. Standardised and non-standardised tests used in the test battery ordered and categorised by cognitive construct.

Cognitive Construct	Test given	Test name	Standa-rdised	Individual/ Group	Test Number	Administration time, items, maximum score	Reliability of measure used
Working memory	Forwards digit recall	Working Memory Test Battery for Children (WMTB-C)	✓	Individual	Individual test number 1	4-minutes per test (approx.) 54 items-maximum score of 54.	Alloway, Gathercole, Willis and Adams (2004) reported a test–retest reliability coefficient of .81 for digit recall.
Working memory	Backwards digit recall	digit recall	✓	Individual	Individual test number 2	4-minutes per test (approx.) 36 items-maximum score of 36	Alloway, et al. (2004) reported that test–retest reliability for children aged between 5 and 8 years was .53
Phonological Processing	Rapid letter naming (Amended by replacing all c with g because of sound in Austria).	Comprehensive Test of Phonological Processing (CTOPP)	✓	Individual	Individual test number 3	1.5-minutes per test (approx.) Two sets of 36 items=total 72. Scored by timed score of total reading.	Manual: For children aged 6, alternate form immediate reliability coefficient was .91
Phonological Processing	RAN digits	Comprehensive Test of Phonological Processing (CTOPP)	✓	Individual	Individual test number 4	1.5-minutes per test (approx.) Two lots of 36 items=total 72. Scored by timed score of total reading	Manual: for children aged 6 alternate form immediate reliability coefficient was .91
Phonological Processing	Phoneme deletion	York Assessment of Reading Comprehension (YARC)	✓	Individual	Individual test number 8	4-minutes (approx. Total tested items 17. Maximum score 17.)	Test reliability, according to the manual was a Cronbach's α of .93. Reliability calculated on the data set yielded a Cronbach's alpha of .80.
Number knowledge	Number reading	Experimental	×	Individual	Individual test number 5	4-minutes (approx.) Total 52 items. Maximum score 52	Item by item, Cronbach's alpha of .92
Visuo-spatial working memory	Forwards block recall	Working Memory Test Battery for Children (WMTB-C)	✓	Individual	Individual test number 6	4-minutes per test (approx.) Total 54 items. Maximum score 54.	The manual reported reliability for forward recall as a test-retest for Years 1 and 2 of .63 (Pickering & Gathercole, 2001). Reliability was calculated for Backwards Block Recall from the first 11 items. The analysis resulted with an alpha of .53 with a warning and including practice trials – see reliability table document for more information
Visuo-spatial working memory	Backwards block recall	Constructed based on the Working Memory Test Battery for Children (WMTB-C) forwards block recall	✓	Individual	Individual test number 7	4-minutes per test (approx.) Total 48 item. Maximum score 48.	
Counting speed	Verbal counting	Experimental	×	Individual	Individual test number 8	60-seconds Scored as highest number correctly counted up to in 1 minute (allowing 1 error).	
Reading	Word reading	Test of Word Reading Efficiency (TOWRE-2), Sight Word Efficiency (SWE)	✓	Individual	Individual test number 9	45-seconds per test Scored as total number of words read correctly in 45 seconds	Test-retest reliability for the Sight word efficiency task in children aged 6-7 on Form B was .93 (Torgesen, Wagner & Rashotte, 2012).
Reading	Non-word reading	Test of Word Reading Efficiency (TOWRE-2),	✓	Individual	Individual test number 10	45-seconds per test Scored as total number of non-words read correctly in 45 seconds	Test-retest reliability for the phonemic decoding efficiency task in children aged 6-7 on

		Phonemic Decoding Efficiency (PDE)					Form B was .93 (Torgesen, Wagner & Rashotte, 2012).
Symbolic and non-symbolic magnitude	Magnitude comparison	Experimental	×	Group	Group test. Day 1 Booklets 1 (4 subtests) and 3 (5 subtest).	30 seconds per test (x3 symbolic, x6 non-symbolic subtests) 48 items subtest 1, 2 and 4 Test 1. 60 items subtest 3, Test 1. 60 items subtest 1, 2 and 3, Test 3. 48 items in subtest 4. Scored as number correct in time limit.	Parallel forms – Reliability of scale .92 (seven subtests into a reliability analysis (total scores) and got Cronbach's alpha of .92)
Number knowledge	Number writing/ transcoding	Experimental	×	Group	Group test. Day 1 Booklet 2 (subtests 1 and 3), Day 2 Booklet 2 (subtests 1 and 5).	4-minutes per test (x4) 13 items per trial. Maximum score 52.	Parallel forms – reliability of scale is .88
Mathematical ability	Numerical Operations	Wechsler Individual Achievement Test (WIAT II)	✓	Group	Group test. Day 1, Booklet 2, subtest 2	15-minutes 15 items. Scored as number correct. Maximum score 15.	Task reliability on items 7-15: Cronbach's alpha of .71 Ran KR-20 and got an r of .71
Non verbal reasoning	Ravens Standard Progressive Matrices	Ravens Standard Progressive Matrices	✓	Group	Group Day 2 Booklet 1	15 minutes. 36 items increasing in difficulty. Score out of 36.	Cronbach's alpha of .75. KR20 also .75
Arithmetic fluency	Addition Subtraction	Adapted from Westwood, P., Harris-Hughes, M., Lucas, G., Nolan, J., & Scrymgeour, K. (1974). One minute addition test - one minute subtraction test. Remedial Education, 9(2), 70-72. And from Goebel et al (2014)	✓	Group	Group Test Day 2 Booklet 2 subtests 2 and 3	One minute addition and subtraction 30 questions of each. Maximum score 30 for addition. 30 for subtraction.	Addition α of .89 and subtraction α of .85 KR20 Addition .89, subtraction .85
Number Knowledge	Number Identification	From Goebel et al. (2014)	×	Group	Group Test Day 2 Booklet 2 subtest 4	8 items, maximum score 8	Reliability on Time 1 dataset on items f-h yielded an α of .82. Reliability on all items (Cronbach's alpha) .67 KR20 was also .67
Symbolic (Numerical Ordinality)	Ordinality task	Experimental	×	Group	Group Day 2 Booklet 2 subtest 6	90 seconds. 80 items. Maximum possible score 80.	
Single digit writing	Number knowledge	Experimental	×	Group	Group Test Day 2 Booklet 2 subtest 7	3 subtests, 9 digits each, total score correct = 27, also scored number of digits mirrored	All items recoded into numbers, Cronbach's alpha .81 Parallel on totals for three tests: .33

Additional materials for group testing

During group testing sessions children were shown a PowerPoint presentation which was used to demonstrate each task to enable the children to have a clear understanding of what was expected for each task.

Additional materials for individual testing

A record form was used for each individual child during individual testing. The record form contained individual record forms for each of the standardised and non-standardised tests used.

A stimulus booklet was used for the tasks which required the child to read from a page. The stimulus booklet consisted of a numerical number line displaying forwards and backwards arrows (used for the forwards and backwards digit recall tasks when children required further help with understanding what was expected of

them in the tasks, although this was rarely used), practice items and pages of letters and numbers (form A and form B – for both the letter and digit RAN tasks), columns of numbers over two pages for the number reading task, 24 coloured images (used as visual prompts for the phoneme deletion task), and lists of practice items and words and non-words (used for the word and non-word reading tasks).

Visuo-spatial working memory was assessed with the Corsi blocks task. The apparatus for this was a set of blocks glued in set positions on a board. The arrangement for the blocks and the sizes were as provided by The Working Memory Test Battery for Children (2001).

The apparatus consisted of nine 3cm square wooden blocks on a wooden board (28.5cm x 23cm). To aid the experimenter, the numbers 1–9 were painted on the side of each block facing away from the child (see Fig. 1a and 1b).

2.1 Transcoding tasks

Number Reading. This task was administered during individual testing. It consisted of 52 numbers in the form of Arabic digits, Calibri (body) font size 20, listed over two A4 pages, with 27 numbers on page one and 25 numbers on page two. Four single-digit numbers, 24 two-digit numbers, 16 three-digit numbers, and eight four-digit numbers were presented. Single digits shown on page one were 1, 4, 6, and 7 (as ordered). Two-digit numbers shown on page one were 16, 70, 25, 68, 56, 91, 48, 27, 79, 13, 30, 47, 11, 80, 73, 42, 34, 81, 32, 89, 53, 15, and 40 (as ordered).

On page two the first number shown was the two-digit number 64, followed by three-digit numbers ordered as 200, 304, 600, 190, 190, 123, 643, 700, 203, 300, 560, 340, 107, 242, and 349. Finally, four-digit numbers were ordered as 8000, 2150, 1015, 2609, 1300, 3791, 1002, and 1060. All four-digit numbers were on the second page.

Number Writing. This task consisted of 52 numbers in the form of Arabic digits. These were exactly the same as the numbers used in the number reading task (see above). This test was administered as part of the group testing and was split across day one and day two. The test was administered as 4x13 item blocks (two sub-tests each day). The first part of the test for each day consisted of 1 single-digit, 9 two-digit, 2 three-digit and one four-digit items and the second part of 1 single-digit, 3 two-digit, 6 three-digit and 3 four-digit items. Number entry was made next to [an easy to identify] illustration, for example, a carrot, a mouse, a tick, a chair, a dog etc. as done in Imbo et al. (2014). Illustrations were displayed in a 1.8x4cm box on the left-hand side of the page. The full list of the numbers for the number writing task in order, together with the accompanying descriptor pictures, and details of the test they were contained in is attached in Appendix 2. A breakdown of the numbers by number type is contained at Appendix 3.

Number Identification. This task consisted of eight rows of numbers from which the participant selected the correct number (numbers are contained in Appendix 4 and the target number is highlighted). The first two questions contained four items to select from, whilst the last four questions contained five items each. One single digit, four two-digit and three three-digit items had to be identified across the task. The non-correct items reflected partial answers e.g. 8 (correct answer 28), inverted answers e.g. 14 (correct answer 41), and syntactic additive errors e.g. 10063 (correct answer 163). This test was taken from Göbel, Watson, Lervåg and Hulme (2014).

2.2 Arithmetic tasks

One Minute Addition and One Minute Subtraction. Each task contained 30 items set out over one page in two columns, typed in Calibri (body) font size 24. All questions were single digit (one to nine) questions presented in a written format e.g. 2 + 1 =, and did not require carrying or borrowing. The questions began with small sums (e.g. 2 + 1 = or 2 – 1 =) and gradually increased in difficulty (e.g. 7 + 6 = or 7 – 6 =).

This test was adapted from Westwood et al. (1974).

Numerical Operations. This test contained 15 questions from the Wechsler Individual Achievement Test 2nd Edition (WIAT-IIUK; Wechsler, 2005) adapted for group use. Adaptations were: number 10 was omitted from the number sequence in question three; and questions seven to 15 were changed from being vertical to horizontal to reflect the way they are typically used in primary school settings. All questions were given in paper format presented as a part of an A4 testing booklet. Questions one to six were dictated by the researcher and required an answer to be written in the correct box (e.g. "put a circle around all the numbers in box one"). Questions one and two required identification of single digit numbers, question three, four, and five required single and double digit number writing and question six required counting the total of an object, with a single digit total (8). Questions seven to 15 were administered as formal arithmetic questions (e.g. 8 + 5 =) and included five addition questions increasing in difficulty, three subtraction questions increasing in difficulty, and one multiplication question. The children had 15 minutes to complete questions seven to 15.

2.3 Magnitude comparison

Magnitude comparison.

These tasks were administered as a part of the group testing during day one and consisted of symbolic and non-symbolic tasks. A total of three digit comparison (symbolic) and six dot comparison (non-symbolic) tasks were administered (two digit and two dot tasks in the booklet administered first and four dots and one digit task in the booklet administered second). Symbolic and non-symbolic tasks were presented across two A5 booklets with six rows of items presented on each page and one pair on each row. Each individual item was presented in a box 25mm by 25mm. The boxes were 19mm apart and 42mm from the left hand margin, and 38mm from the right hand margin. An instruction page with an example task was given prior to each task. Booklet one contained two symbolic tasks and two non-symbolic tasks. The first exercise test was a practice comparing digits (symbolic), preceded by two worked through examples (see below for further details). These two practice subtests contained 48 items. The second exercise was a practice test comparing small squares (non-symbolic comparison) and was also preceded by two worked examples. This was followed by exercise 'Number 1' (symbolic) and exercise 'Number 2' (non-symbolic), both preceded by two worked examples.

Booklet two firstly contained exercise 'Number 3', a non-symbolic comparison task, preceded by two worked examples. It then contained exercise 'Number 4', a symbolic (digit) comparison task, preceded by two worked examples, followed by exercises 'Number 5', 'Number 6' and 'Number 7', all non-symbolic comparison tasks, with two preceding worked examples. More information on the items (average and SD of problem size, distance, number range and ratio) for these tasks can be found in Appendix 5. The tests were taken from Göbel, Watson, Lervåg and Hulme (2014). The booklets were designed so that the children couldn't see the test page until told to turn over, once the time had started. Coloured tabs were used to help find the correct pages (these were the exercise number page and the example page, and had no assessment data on).

Item Design

Symbolic magnitude comparison. Arabic digit pairs, Calibri, font size 48, consisted of numbers in the range of one to nine. Pairs of digits were designed to be 'close' and had a numerical distance from each other of one to four, or 'far' from each other and had a numerical distance of five to nine. The symbolic comparison tests were Practice 1 (P1), Exercise Number 1 (E1) and Exercise Number 4 (E4). In the Practice 1, 48 items were presented. The pairs of these number distances were mixed, with and average problem size of 9.71, and a range of 14.

In the two proceeding tasks 60 items were presented, one task presented numbers in the close distance and one task presented numbers in the far distance, both with an average size of 10 (range=5). In task one the

digits displayed consisted of Arabic symbols one to nine (far) and in task two digits consisted of Arabic symbols three to seven (close).

Non-symbolic comparison. Displays of dots presented in this task ranged from five to 40. In the practice task (P2) 48 items were presented and number ratios were mixed and display size varied from consisted of five to 13 dots. For three of the non-symbolic comparison tasks there were 48 items (E2, E6, E7), pairs of dots displays were matched on surface area (SA), i.e. the overall amount of black was the same in both displays for each pair. These three subtests varied on the ratio between the pairs (E2: 7:8, E6: 5:6, E7: 3:4). The remaining two tasks (E3, E5) had 60 items each, one with a close distance and one with a far distance and both with a problem size of 15 and all dots in E3 and E5 were of the same size (SS).

2.4 Ordinality task

Digit Ordinality. There were 80 items. Each item consisted of three Arabic digits (Arial, font size 48) between 1 and 9, each digit was presented in a box 2.86cm by 2.22cm. The three digits were either in ascending order (e.g., 2-4-6) or not in order (e.g., 2-6-4). On each page there were 10 items set out in two columns with each column containing five items. In total there were 35 ascending triplets, their number varied between 2 and 6 on each page (see Appendix 6). An arrow pointing downwards was set out to the left of each column to indicate the way in which the items should be completed. Numerical distance between the three sets of digits in the ascending condition was either one (e.g. 1-2-3), two (e.g., 1-3-5) or three (e.g., 1-4-7). The numerical distance of ascending triplets was not equally distributed across the pages.

2.5 Timed counting task

The researcher used a marking sheet which contained two squares of numbers to record counting errors. One square presented numbers one to 100 and the second square presented 101 to 200. The two number squares were set out over two pages and each number (Comic Sans, font size 12) was presented in a box 1.6cm by 1.65cm. The participant could not see the sheet.

2.6 Domain-general tasks

Nonverbal IQ. The Ravens Standard Progressive Matrices Test was used to test non-verbal IQ. Items were set out in an A5 booklet containing 36 items which increased in difficulty. Items A1 and A2 were used as practice items. The remainder of items presented were A3 to A12, B1 to B12, and C1 to C12. This was adapted for group use by including the answer options on the page.

Single Digit Writing. An A4 page consisting of nine joined boxes (each box 2cm by 2cm) set out horizontally approximately 10cm from the top of the page was presented. This task was repeated three times consecutively. The digits were read out in three blocks of nine digits: Block 1: 5 2 8 3 1 9 6 4 7; Block 2: 4 1 7 3 8 2 9 5 6; and Block 3: 8 9 4 6 5 7 1 3 2.

Digit Recall Forwards and Backwards. The Working Memory Test Battery for Children (WMTB-C, 2001) was used. For each task, the researcher used a marking sheet set out on an A4 page. The marking sheet consisted of practice items (three for forwards digit recall and two for backwards digit recall) and test items (54 individual test items for forwards digit recall and 42 individual test items for backwards digit recall).

Block Recall Forwards and Backwards. For each task, the researcher used a marking sheet set out on an A4 page. The marking sheet consisted of practice items (three for forwards block recall and two for backwards block recall) and test items (54 individual test items for forwards block recall and 24 individual test items for backwards block recall). The CORSI block arrangement from the Working Memory Test

Battery for Children (WMTB-C, 2001) was used (see Figure 1a and 1b). This consisted of 9 individual wooden or plastic blocks set out on an overall platform 27.8cm width by 22.7cm depth. Each of the nine individual blocks was 3cm by 3cm. The individual blocks were numbered (1-9) on one side, to aid the experimenter (these numbers were not shown to the children). The original test did not include a backwards version, so the Graz team designed this based on the items of the forwards version. Block Recall Backwards includes the same items as Block Recall forwards, but in a different order.

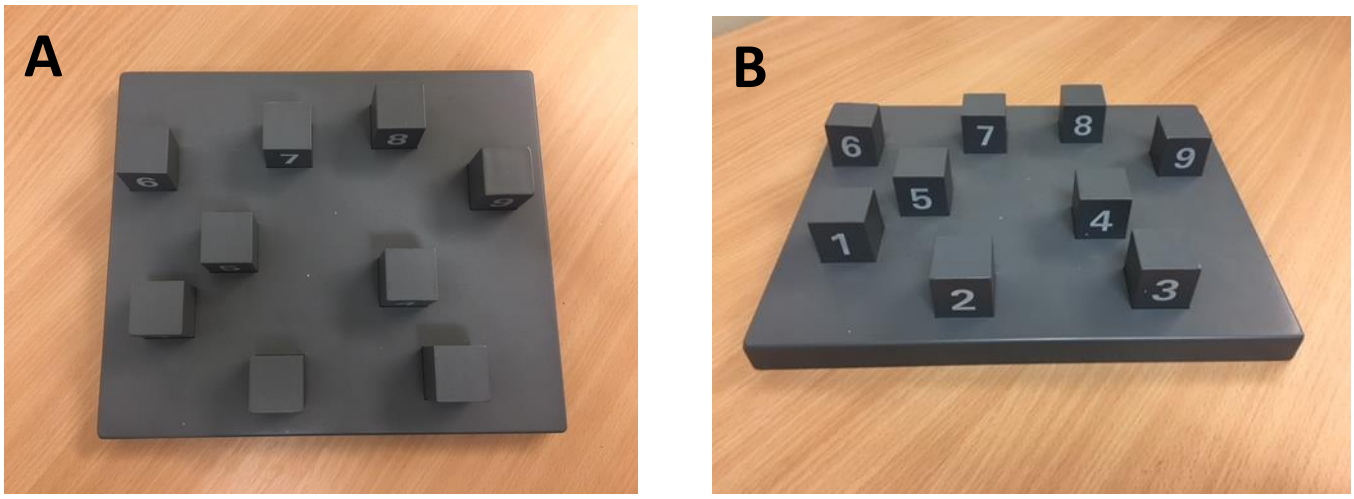

Figure 1. Photograph of the Corsi blocks used; A: from above, B: experimenter view

Rapid letter and digit naming. For these tasks the Comprehensive Test of Phonological Processing 2 (CTOPP 2, 2013) was used as a part of individual testing. For each test the researcher used the CTOPP marking sheet which consisted of the administration instructions and two blocks of either letters or numbers (depending on the test). The two blocks were labelled form A and form B.

In rapid letter naming, the stimuli presented to the child consisted of 36 letters in each block, and were presented on an A4 page. For rapid letter naming the letters were typed in BonnBold size 46. The letter c was replaced by g as the team in Graz, Austria suggested that children in Austria do not routinely see c as a sole letter. This change was made in the UK too to ensure the same materials were used across both testing sites.

The stimuli for rapid number naming consisted on 36 numbers (1-9) in each block, presented on an A4 sheet. The material was taken directly from the CTOPP 2 (2013) manual.

3. Procedure

Children took part in two one-hour group testing sessions (one on the first day and one on the second day of testing) and one 30-minute individual testing session. Group testing sessions were held in the children's classrooms and individual testing was conducted in a different part of the school on a one to one basis. The first group testing session took place on the first day of testing, always prior to any individual testing. Some children were then tested individually following the first group testing session (i.e. on day one) and some children were tested individually following both group testing sessions (i.e. on day two). The order of tasks was the same for all children. Individual testing sessions comprised of 10 individual tests and group testing sessions comprised of nine individual tests (see Table 1).

Group testing sessions

Prior to group testing, children were told that there were three rules they had to follow: try your best, wait for the green light (when a green traffic light was displayed on the screen before they were allowed to start) and to stop and put their hands up as soon as the researcher said stop. A PowerPoint presentation was displayed for each group testing session to support children's understanding of the tasks they were being asked to complete.

3.1 Day 1 Testing

Booklet order:

Day 1 Booklet 1: Symbolic and non-symbolic magnitude comparison (Part 1)

Day 1 Booklet 2: Number writing (Part 1), numerical operations, number writing (Part 2).

Day 1 Booklet 3: Symbolic and non-symbolic magnitude comparison (Part 2)

Day 1 Booklets 1 and 3

Symbolic and non-symbolic magnitude tests

Participants were given A5 booklets containing a total of nine magnitude comparison tasks. Prior to each task children were shown an example on the PowerPoint presentation which corresponded to the practice example in their booklets. For each pair, children were asked to tick the box containing the bigger number or the larger number of dots, depending on the task presented. Children were told not to count the dots.

Children were given 30 seconds to complete as many comparisons as possible. If any child started before the time started or continued after the time had stopped the items completed were crossed out by the researcher. This happened rarely. On completion of one task, children were then asked to turn over the pages by finding the correct coloured tab, until they got to the next comparison task header page and examples page, where the researcher explained the next comparison task. These pages contained no tested materials.

Coding and scoring

Children were given one point for each item in which they correctly ticked the larger number in the pair (symbolic) or the larger number of dots in the pair (non-symbolic). The maximum score for each of the practice tasks was 48. The maximum score total score for Exercise 1 was 60, 48 for Exercise 2, 60 for Exercises 3-5, 48 for Exercises 6 and 7.

Day 1 Booklet 2

Children were provided with an A4 booklet containing two number writing (transcoding) tasks and a numerical operations task, taken from Wechsler Individual Achievement Test 2nd Edition (WIAT II).

Number writing

Children were instructed to write the Arabic forms of dictated numbers next to a given illustration. For example, the researcher would say "write the number six next to the chair". Each number was spoken once but could be repeated a further two times, on request. This task was not timed.

Coding and scoring

Children were given one point for each number written correctly. If the child mirrored the digits, but the overall answer was correct, this was noted but that answer was marked as correct. The maximum possible score achievable was 52 (13x4).

Numerical operations

Wechsler Individual Achievement Test 2nd Edition (WIAT II) (adapted)

The Numerical Operations subtest was presented as a paper-and-pencil test used to measure numerical ability. The first part of the task contained six items and was dictated by the researcher. The tasks involved identification of single digit numbers (questions one and two), number writing single and double digit numbers (questions three to five) and counting (question six). The second part of the task contained nine items and children were then given up to 15 minutes to complete as many as they could. Items consisted of 5 addition items (including two double digit items), 3 subtraction calculations (including one double digit item) and one multiplication item. These were set out horizontally as the norm with this age group, rather than using formal columnar methods. Children were also reminded by the researcher that they could use their fingers to help solve these. A stopwatch was used for timing. If children finished before the time given they were asked to draw a picture to ensure they did not disturb those still working. At the end of time children were told to stop and raise their hands.

Coding and scoring

One point was given for each correct item. If the child provided the correct answer but with mirrored digits the answer was marked as correct. The maximum possible score was 15. In addition, for each child we also calculated the number of correct items on this test excluding the first six items (t1nocore, maximum score = 9). The first six items are number writing and transcoding items, while items from item 7 onwards are arithmetic items.

3.2 Day 2 Testing

Booklet order:

Day 2 Booklet 1*: Nonverbal IQ– Ravens Standard Progressive Matrices*

Day 2 Booklet 2: Number writing (Part 3), one minute addition, one minute subtraction, number identification, number writing (Part 4), ordinality (digits only), single digit writing

Day 2 Booklet 1

Ravens Standard Progressive Matrices

Participants were given one A5 booklet containing 36 Ravens Matrices items, which got progressively more difficult. Children were told they had to find the missing jigsaw piece and were asked to put a tick over the box containing the missing pattern piece. Prior to the task, children were shown an example on the PowerPoint presentation which corresponded to the practice example in their booklets and this example was completed as a group. A further example was also completed, using the PowerPoint presentation and feedback was provided. Children were then allowed 15 minutes to complete the booklet. If any child started before the time started, or continued after the time had stopped, the items completed in the extra time were crossed out by the researcher. A stopwatch was used for timing. At the end of time children were told to stop and raise their hands.

Coding and scoring

One point was given for each correct item. The maximum possible score was 36.

Day 2 Booklet 2

Children were provided with an A4 booklet containing the second part of the number writing (transcoding) task (as detailed above). The day two booklet also contained the following tasks:

One minute addition and subtraction

For this task children were asked to complete as many of the 30 addition questions and 30 subtraction questions as they could within 60s for each subtest. The questions were displayed in two columns, going down the page. The questions were set out horizontally (not columnar) and got progressively more difficult. If any child started before the time started or continued after the time had stopped the items completed were crossed out by the researcher. A stopwatch was used for timing. At the end of time children were told to stop and raise their hands.

Prior to task completion children were shown an example on the PowerPoint presentation which corresponded to the practice example in their booklets and this example was completed as a group. Children then had one minute to answer as many one digit addition calculation questions (for which the answer could cross the ten boundary) as they could.

This process was repeated for one minute subtraction (questions included some double digit minuends with all single digit subtrahends and difference). There were 30 available questions.

Coding and scoring

One point was given for each correct item. If the child provided the correct answer but with mirrored digits, this was marked as correct. The total number of correct items was calculated for the one minute addition and the one minute subtraction task separately.

Number Identification

Children were presented with eight number identification items. The researcher said a number and the children were required to select it from a row of items including 3-4 distractor items. For example, the researcher would say “put a circle round the number twenty five”. Children then had to choose between five possible answers which were; 502, 5, 25, 50, and 52. This task was untimed, but paced by the experimenter.

Coding and scoring

One point was given for each correct item identified. If a child circled two or more numbers, they were given a score of 0 for that item. The maximum score possible was 8.

Ordinality (Digits only)

Children were presented with eighty rows, with 10 rows presented on each page, each containing three single digit numbers. Items were set out in two columns: Column A and Column B with five items under each column on one page. Children were asked to tick the row of three numbers if the numbers were increasing in numerical size, and to draw a line through the row if the numbers were not ordered by numerical size. Prior to the task children were shown an example on the PowerPoint presentation which corresponded to the example in their booklets and this example was discussed and explained with the group. Children then completed six rows as a completion example and these were discussed to ensure all children understood the task.

Children were then given 90 seconds to tick or cross as many of the eighty rows as they could, using the rule. Children were instructed to work down the columns when assessing the rows. If any child began the task before the time started, or continued after the time had stopped, the items completed during the extra time were crossed out by the researcher. A stopwatch was used for timing. At the end of time children were told to stop and raise their hands.

Coding and scoring

One point was given for each item correctly identified as increasing in numerical size. The possible maximum score for this task was 80.

Single Digit Writing

In this task, children heard a series of 9 single digit numbers dictated separately by the researcher and wrote it in a grid which corresponded to the grid displayed on the PowerPoint presentation. There were nine boxes in the grid and children entered numbers in the boxes from left to right. The numbers represented the single digits 1-9 and were repeated three times, over three separate tasks, each time in a different order. This was an untimed task and the next number was not said until all children had written the previous number.

Coding and scoring

One point was given for each correct item. Errors which contained numbers which had been mirrored were also recorded. The maximum score for this task was 27.

3.3 Individual testing session

At the start of individual testing the researcher introduced themselves to the children and asked the child for their name. After the children had given their name they were informed that the session would be recorded so the researcher was able to listen back to it, and they were told not to worry about being recorded and just to do the best they could do. Recorders were then switched on and the researcher said the time, date, and their own initials. Children were then asked to draw a picture of a smiley face on the front of the recording booklet, this was done so the researcher could record the child's handedness.

Digit recall forwards

Working Memory Test Battery for Children (WMTB-C)

For this task children were asked to repeat, in the same order, a number string presented verbally by the researcher. Each session began with three practice trials following administration instructions. Children began testing on the span they were able to complete during practice trials, which started at a span of one up to a span of three. Providing that children completed the first span successfully (four correct trials) they moved on to the next span until they were unable to get more than three trials correct in any one span. On each span up to five number strings could be presented.

Coding and scoring

Following the scoring instructions from the test manual, the span for each child was recorded as the highest span in which children correctly recalled four or more items. The number of correct items was also recorded. If a child successfully repeated the number sequences on the first four trials for a given span a total of six points was given for that given span. The full six points were also given for any span not administered due to successful completion of the practice trials. For example, if a child began the test items on span three, due to successful completion of the practice trials up to and including span three, six points were given for one span trials and six points were given for two span trials. Maximum possible score for this task was a span score of 9 and an item score of 54.

Digit recall backwards

Working Memory Test Battery for Children (WMTB-C)

Children were told that the researcher would say a series of numbers and that the child had to listen, wait till the experiment had finished speaking and then repeat the numbers in a backwards order. For example, if the

experimenter said the numbers 1, 6, 8, the child would have to say the numbers as 8, 6, 1. Children were given practice trials one and two which consisted of two numbers. If the child correctly repeated the numbers in a backwards order, span two was administered before completing practice items three and four, both of which contained three digits. Children then continued with span three as per the administration instructions. If a child was unable to understand the procedure of repeating the numbers in a backward order the child was shown a number line to support their understanding or repeating numbers backwards and the administration instructions were followed, as set out in the test manual. Children could continue to the next span if they correctly recalled the backwards string of four items in a given span. Testing discontinued following three incorrect trials in a given span.

Coding and scoring

Scoring and coding followed the same rules as the forwards digit recall test (as described above). Maximum score on this task was a span score of 7 and an item score of 36.

Rapid letter naming

Comprehensive Test of Phonological Processing (CTOPP)

Children were first shown the practice items consisting of 6 letters which the child was asked to say/sound out. Once the child had said the letters/letter sounds the child was shown the page of letters (form A) and was shown the direction in which the letters should be said (left to right). Children were timed from the start to completing reading the letters on the page. The same procedure was then repeated for form B.

Coding and scoring

The total number of items read correctly were recorded and errors and omissions were recorded. The time, in seconds and milliseconds, in which the child completed reading the items was also recorded.

Rapid digit naming

Comprehensive Test of Phonological Processing (CTOPP)

Rapid digit naming followed the same procedure as rapid letter naming.

Coding and scoring

The total number of items read correctly were recorded and errors and omissions were recorded. The time, in seconds and milliseconds, in which the child completed reading the items was also recorded.

Number reading

Children were shown a list of 52 written numbers in Arabic digit format starting from single digits going up to multi-digits (up to thousands) i.e. increasing in complexity. In total four single digits, 24 two-digit, 16 three-digit, and eight four-digit numbers were presented. These numbers were the same as those presented in the transcoding task. However, here they were ordered by size. Children were shown the list of numbers which were presented in the stimuli booklets over two A4 sheets. A piece of card was used to cover all the numbers and the card was moved down after each number was read allowing the child to see the next number. All children read out all items up to the first four-digit number. If more than three three-digit numbers had been read correctly the child moved on to the remaining four-digit numbers. If a child failed to read out three or more three-digit numbers correctly testing was discontinued at this point. This occurred in 60 cases. In exceptional circumstances where children were struggling with two-digit numbers the researcher stopped testing earlier at their own discretion (this happened for 10 children).

Coding and scoring

The researcher transcribed children's utterances on a separate marking sheet during administration of the task. Children's answers were also audio-recorded. One point was given for each number read correctly. Errors were also recorded. The maximum possible score was 52.

Block recall forwards

Working Memory Test Battery for Children (WMTB-C)

In the block recall task children were shown the Corsi blocks (see Figure 1). Children were told that the researcher would tap on the blocks and that the child had to tap the same blocks in the same order. Each session began with three practice trials following administration instructions. Children began on the span they were able to complete successfully during practice trials. Providing that children completed the first span successfully (four correct trials) they moved on to the next span until they were unable to get more than three trials correct in any one span. Testing discontinued once a child had incorrectly recalled three block recalls for any given span.

Coding and scoring

Following the scoring instructions from the test manual, the span for each child was recorded as the highest span in which children correctly recalled four or more sequences. The number of correct trials was also recorded. If a child has successfully repeated the number sequences on the first four trials for a given span a total of six points was given for that given span. The full six points was also given for any span not administered due to successful completion of the practice trials. For example, if a child began the test items on span three, due to successful completion of the practice trials up to and including span three, six points were given for one span trials and six points were given for two span trials. The maximum span score was 9 and the maximum possible item score was 54.

Block recall backwards

Children were told that the researcher would tap the blocks in a forwards order and the child then had to tap the same blocks in a backwards order. In the first instance children were given practice trials one and two; each of those trials consisted of a sequence of two blocks. If the child correctly repeated the blocks in a backwards order span two was administered before completing practice items three and four, both of which contained three blocks. Children then continued with span three as per the administration instructions. If a child was unable to understand the procedure of repeating the blocks in a backwards order the procedure was demonstrated to the child.

Coding and scoring

Scoring and coding followed the same procedure for forwards block recall. The maximum span score was 9 and the maximum possible item score was 54.

Timed counting

Children were asked to count as high as they could within one minute starting at 1. Children were asked to count as quickly, and say the number as clearly, as they could and to stop when the researcher said stop.

Coding and scoring

The highest number reached within 60-seconds was recorded for each child. If the child made one error they were given the score of the highest number they got up to at the 60-second mark. If a child made more than one error the last number the child said correctly was recorded along with the time (in seconds) the second error was made. One error counted as one number omitted or one number said incorrectly. We also calculated the numbers counted per second for each child.

Phoneme deletion

York Assessment of Reading Comprehension (YARC)

Children were shown an image in the stimulus booklet and the researcher said a word and asked the child to repeat the word. Once the child had repeated the word the child was then asked to say the word again with part of the word or a phoneme missing. For example, the researcher would say "say goat" [reply from child], "now say it again without the 'g'". Two practice trials were given for each deletion pattern. 17 trials could be administered but if the child responded incorrectly on more than five consecutive occasions testing was stopped at item 13, otherwise the extra 4 extension items were also administered.

Coding and scoring

One point was given for each correct item. The maximum possible score was 17.

Reading: Sight word efficiency

Test of Word Reading Efficiency (TOWRE-2)

Children were first shown eight practice items of real words which the researcher asked the child to read. If any errors were made on the practice items the researcher corrected the child. The child was then told that they would be given a list of words and they need to read as many words, as quickly and as accurately, as possible within 45-seconds. The experimenter then turned over the page in the stimuli booklet and asked the child to read down the columns and then asked the child to start reading the words. The number of words correctly read was recorded.

Coding and scoring

The total number of words read correctly was recorded. The maximum possible score was 108.

Reading: Phonemic decoding efficiency

Test of Word Reading Efficiency (TOWRE-2)

Children were asked to read a list of practice items which consisted of eight non-words. The rest of the task followed the same procedure as the Sight Word Reading Efficiency, except that the words were non-words.

Coding and scoring

The total number of words read correctly was recorded. The maximum possible score was 66.

4. Reference list for tests and test manuals

Wechsler, D. (2005). Wechsler individual achievement test (2nd UK ed.). London: Pearson Clinical and Talent Assessment.

Pickering, S.J., & Gathercole, S.E. (2001). Working Memory Test Battery for Children. Psychological Corporation Europe.

Raven, J. C., Styles, I., & Raven, M. A. (1998). Raven's Progressive Matrices: SPM plus test booklet. Oxford, England: Oxford Psychologists Press/San Antonio, TX: The Psychological Corporation.

Snowling, M.J., Stothard S.E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., Nation, K., & Hulme, C. (2009). York Assessment of Reading for Comprehension. London: GL Assessment

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). Test of word reading efficiency, second edition. Austin, TX: PRO-ED, Inc.

Wagner R. K., Torgesen J. K., Rashotte C. A., Pearson N. A. (2013). Comprehensive Test of Phonological Processing–Second edition. Austin, TX: PRO-ED.

Westwood, P., Harris-Hughes, M., Lucas, G., Nolan, J., & Scrymgeour, K. (1974). One minute addition test - one minute subtraction test. *Remedial Education, 9*(2), 70-72.

Appendix 1: English Indices of Deprivation

(from: Department for Communities and Local Government. (2015). English indices of deprivation 2015. https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015)

Deprivation 'deciles' are published alongside ranks. Deciles are calculated by ranking the 32,844 neighbourhoods in England from most deprived to least deprived and dividing them into 10 equal groups. These range from the most deprived 10% of neighbourhoods nationally to the least deprived 10% of neighbourhoods nationally, as shown in the table below:

Decile	Decile description	Ranks
1	10% most deprived	1 to 3,284
2	10% to 20%	3,285 to 6,568
3	20% to 30%	6,569 to 9,853
4	30% to 40%	9,854 to 13,137
5	40% to 50%	13,138 to 16,422
6	50% to 60%	16,423 to 19,706
7	60% to 70%	19,707 to 22,990
8	70% to 80%	22,991 to 26,275
9	80% to 90%	26,276 to 29,559
10	10% least deprived	29,560 to 32,844

Appendix 2: Transcoding

Day 1, Part 1

Accompanying picture	Number
Candle	1
Toothbrush	11
Shoes	80
Paint Splat	73
Car	42
Ice-cream	34
Triangle	81
Plane	32
Pencil	89
Bed	53
Ball	700
Hat	203
Square	1300

Day 1, Part 2

Accompanying picture	Number
Leaf	7
Phone	15
Snail	40
Books	64
Chocolate	300
Spoon	560
Apple	340
Bowl	107
Rubber	242
Lipstick	349
Hammer	3791
Ruler	1002
Flag	1060

Day 2, Part 3

Accompanying picture	Number
Flower	4
Cat	16
Cake	70
Dog	25
Cow	68
Fish	56
Moon	91
Smiley	48
Tick	27
Cross	79
Star	200
Tree	304
Bath	8000

Day 2, Part 4

Picture	Number
Chair	6
Table	13
Lollypop	30
Sun	47
Cloud	600
Pig	190
Mouse	220
House	109
Bike	123
Window	643
Carrot	2150
Door	1015
Gloves	2609

Appendix 3: Transcoding items by number type

One and two digit numbers		
Number of items	Symbol	Numbers
4	X	4 6 1 7
4	Teens	16 13 11 15
4	X0	70 30 80 40
10	XX	25 68 56 91 48 27 79 73 42 33
Three digit numbers		
4	X00	200 600 700 300
10	XX0	190 150 220 560 170 240 160 280 390 340
10	X0X	304 109 101 203 107 207 406 201 602 105
10	XXX	198 142 123 643 413 242 266 951 756 349
Four digit numbers		
1	X000	8000
1	XX00	1300
1	XXX0	2150
1	XXXX	3791
1	X0XX	1015
1	X00X	1002
2	XX0X	2609 4701
1	X0X0	1060

Appendix 4: Number identification

(Correct target numbers are highlighted.)

a	8	6	3	9	
b	1	41	4	14	
c	82	28	208	8	20
d	502	5	25	50	52
e	76	17	6	706	67
f	25	235	20035	23	253
g	13	10063	136	15	163
h	472	427	47	42	40027

Appendix 5: Magnitude comparison

		Practice1	Practice2	Exercise1	Exercise2	Exercise3	Exercise4	Exercise5	Exercise6	Exercise7
		Digits	**Dots**	**Digits**	**Dots**	**Dots**	**Digits**	**Dots**	**Dots**	**Dots**
			SS		**SA**	**SS**		**SS**	**SA**	**SA**
					7:8				**5:6**	**3:4**
Number of items		48	48	60	48	60	60	60	48	48
Number of dots/digits		1 to 9	5 to 13	1 to 9	20 to 34	7 to 11	3 to 7	5 to 13	20 to 35	20 to 40
	average	4.85	8.86	5.00	27.05	9.00	4.98	9.00	27.77	29.43
	sd	2.48	2.47	3.19	3.77	1.28	1.26	3.19	4.27	5.51
Distance		1 to 8	1 to 8	5 to 7	2 to 4	1 to 2	1 to 2	5 to 7	3 to 6	6 to 10
	average	3.29	3.27	5.97	3.48	1.43	1.43	5.97	4.96	8.27
	sd	2.04	2.05	0.76	0.65	0.50	0.50	0.76	0.90	1.20
		Mixed	**Mixed**	**Far**	**Mixed**	**Close**	**Close**	**Far**	**Far**	**Far**
Ratio	***min***	0.11	0.38	0.13	0.86	0.78	0.60	0.42	0.82	0.73
	max	0.89	0.92	0.44	0.92	0.91	1.67	0.62	0.88	0.79
	average	0.50	0.70	0.24	0.88	0.85	0.78	0.50	0.84	0.75
	sd	0.24	0.16	0.11	0.02	0.05	0.17	0.06	0.02	0.02
Problem Size	***min***	3	11	7	43	15	7	15	44	47
	max	17	25	13	64	21	13	21	66	70
	average	9.71	17.73	10.00	54.10	18.00	9.97	18.00	55.54	58.85
	sd	3.09	3.09	2.06	6.69	2.06	2.02	2.06	6.92	7.18

Appendix 6: Ordinality (digits)

Column 1				Column 2			
trial number				**trial number**			
practice	4	1	7	practice	3	6	9
practice	9	3	6	practice	4	6	5
practice	2	5	8	practice	9	5	7
1	3	4	5	6	5	7	3
2	3	9	6	7	3	5	1
3	1	3	5	8	6	7	8
4	8	6	7	9	2	5	8
5	4	1	7	10	4	2	3
11	7	1	4	16	5	7	9
12	9	3	6	17	1	4	7
13	8	5	2	18	2	3	4
14	4	6	5	19	6	9	3
15	1	7	4	20	4	6	8
21	2	4	3	26	7	1	4
22	5	2	8	27	3	6	9
23	1	3	5	28	4	2	6
24	5	8	2	29	3	6	9
25	5	7	9	30	3	7	5
31	4	3	5	36	5	6	7
32	4	6	8	37	4	8	6
33	3	4	5	38	4	5	6
34	1	5	3	39	5	6	7
35	3	5	7	40	2	8	5

Column 1				Column 2			
Item number				Item number			
41	1	4	7	46	5	7	6
42	6	7	5	47	2	4	6
43	4	5	6	48	1	4	7
44	8	4	6	49	2	3	4
45	7	4	1	50	5	6	4
51	3	6	9	56	5	3	4
52	7	6	8	57	3	5	7
53	5	9	7	58	2	5	8
54	6	2	4	59	2	4	6
55	6	7	8	60	9	5	7
61	7	6	8	66	9	3	6
62	2	3	4	67	2	5	8
63	4	1	7	68	4	3	5
64	4	2	3	69	4	6	5
65	2	8	5	70	7	1	4
71	5	2	8	76	4	6	8
72	6	7	8	77	5	7	9
73	2	4	6	78	4	8	6
74	5	3	4	79	6	7	5
75	7	1	4	80	3	5	7