Methods of Timepoint 3 (2019)

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

As shown in Table 1, 257 children took part at T3, of which 232 had complete data. The children all attended Year 3 of UK Primary Schools. Of those with complete data, 217 also had complete data from T2, and 195 from T1 and T2. The children at T3 were between seven and eight years old (M_{age}= 97.72 months, *SD*= 3.84, *Min* = 96, *Max* = 106). Of the 232 children with complete data, 115 (49.6%) were males and 117 (50.4%) females. The attrition of children from T1 to T3 by school is also shown in Table 1.

English as an additional language. From the sample tested, seven (3%) children were classed as having English as an additional language (EAL) or as bilingual by the school records. Six of these children also participated at T1 and T2.

Schools Profile. Children were recruited from both suburban and rural primary schools in areas across North and West Yorkshire. Descriptions of the schools' ethnicity, demographics and school area can be found in Table 2. This was the final year of a three-year longitudinal study in North and West Yorkshire. Testing took place between May and July 2019 in ten North and West Yorkshire Primary Schools. There was one school less than at T1 and T2, as one of these T1/T2 schools was an infant school, and it was not possible to obtain consent from the head teacher of the junior feeder school that most children now attended. This school is School 5.

Special Educational Needs. Children with developmental disorders and/or neurological disorders were not excluded from the study; however around five children in the schools did not take part, at their teachers request where the child had significant developmental delay meaning they would not be able to assess the test at all. Twenty-three of the children with complete data were included on their school SEN register. As a comparison, in England nationally, 12.6% of children attending primary school had special educational needs as of January 2019 (National Statistics, 2019).

Table 1: Attrition at Time 3

School No.	Total No. of Children taking part T1	No of Children with complete data at T1	Total No. of Children taking part at T2	Total No. of children with complete data T2	No. of children left at T2	Children with complete data at T1 and T2	No. of children joined at T2	% Attrition at T2	Total No. of children taking part in T3	No. of children T3 with complete data	No. of children left T3	Children with complete data at T1, T2, T3	No. of children joined at T3	% attrition at T1 to T3	% attrition T2 to T3
1	24	23	28	27	4	20	6	16.67	22	19	6	16	0	25.00	22.22
2	41	41	38	37	3	35	2	7.32	41	33	3	30	4	14.63	7.89
3	20	19	20	20	1	18	1	5.00	20	19	0	17	1	5.00	0.00
4	32	32	36	35	3	29	7	9.38	34	32	7	21	6	28.13	19.44
5	56	56	32	32	24	32	0	42.86	0	0	32	0	0	100	100.00
6	20	20	20	20	0	20	0	0.00	17	16	3	16	0	15.00	15.00
7	8	7	10	10	0	7	2	0.00	10	10	1	6	1	12.50	10.00
8	26	23	28	27	3	20	5	11.54	27	23	4	16	1	26.92	14.29
9	29	29	31	31	0	29	2	0.00	29	26	3	26	1	10.34	9.68
10	19	18	22	22	0	18	3	0.00	22	20	0	18	0	0	0.00
11	39	38	36	36	4	34	1	10.26	35	32	6	29	3	27.03	16.67
Totals	**314**	**306**	**301**	**297**	**42**	**262**	**29**	**13.38**	**257**	**230**	**65**	**195**	**17**	**22.61**	**21.59**

Table 2: Description of Schools

School No.	School Type
1	Academy Coeducational Primary School. 12% SEN, 8.9% EAL, 25% FSM recorded 2018/2019
2	Academy Coeducational Primary School 20.3% SEN, 1.6% EAL, 12.2% FSM recorded 2018/2019
3	Church of England Voluntary Controlled Coeducational Primary School 13.8% SEN, 0.6% EAL, 6.9% FSM recorded 2018/2019
4	Private non-selective coeducational school No governmental data available
5	*Academy Coeducational Primary. DID NOT TAKE PART AT T3.*
6	Coeducational Community Primary School 10.4% SEN, 2.5% EAL, 6% FSM recorded 2018/2019
7	Church of England Voluntary Controlled Coeducational Primary. 9% SEN, 1.3% EAL, 10.3% FSM recorded 2018/2019
8	Academy Coeducational Primary School 11.9% SEN, 5% EAL, 22.2% FSM recorded 2018/2019
9	Coeducational Community Primary School 8.7% SEN, 8.3% EAL, 5.3% FSM recorded 2018/2019
10	Coeducational Community Primary School 9.6% SEN,3.8% EAL, 9.6% FSM recorded 2018/2019
11	Coeducational Community Primary School 2.2% SEN, 3.3% EAL, 1.8% FSM recorded 2018/2019

Notes: School characteristics data from compare-schools-performance service.gov.uk (new service since T2).

Key: SEN=Special Educational Needs (National average in 2019 12.6%,) EAL= English as an additional language (National average in 2018/2019 21.2%) FSM=Free school meals eligibility last six years (National Average in 2019 23%) (National Statistics, 2019).

Overview of Schools. Ten schools took part at T3. Five of the schools were in an urban/city location, one was a coastal town and one a rural hamlet. The final three schools were in a rural town fringe location. Three of the schools were academy coeducational primary schools, four schools were coeducational community primary schools, one was a private (non-state controlled) school, taking children aged two to eighteen, which was coeducational until the age of eleven; the final two schools were Church of England Voluntary Controlled Coeducational Primary Schools.

Ethics. The project had ethical approval from the University of York, updated for Time 3 methods. Fully informed written consent was obtained from head teachers at participating schools. Informed consent was obtained by parental opt-out at eight schools and by opt-in consent at two schools.

The study was approved by the University of York Psychology Department Ethics Committee. (Reference number 559), updated for Time 3. 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 in eight schools, with the exception of two schools who chose an opt-in approach. School 11 in the study had used opt-in consent also in previous years. However, school 1 changed to opt-in at this time point only, at the class teacher's request.

2. Materials and Stimuli

Materials used consisted of standardised and non-standardised tasks. Tasks measured a number of cognitive constructs and are detailed in Table 3 for tests completed during group testing and in Table 4 for tests completed individually, in the order they were completed.

Table 3: Group-administered testing: Tests used in the test battery at Time 3

Construct	Test	Standardised	Test Day & Number	Administration Time, Items, Maximum Score	Reliability at T3
Symbolic and Non-symbolic magnitude comparison	Magnitude Comparison	Experimental	Day 1, Tests 1 and 3 Day 1 Test 1 Contained: Practice 1 (symbolic) Practice 2 (non-symbolic) Exercise 1 (symbolic) Exercise 2 (non-symbolic) Day 1 Test 3 Contained: Exercise 3 (non-symbolic) Exercise 4 (symbolic) Exercise 5 (non-symbolic) Exercise 6 (non-symbolic) Exercise 7 (non-symbolic)	30 seconds per test (3 symbolic, 6 non-symbolic) Practice 1: 30 seconds. 48 items. Maximum score 48. Practice 2: 30 seconds. 48 items. Maximum score 48. Exercise 1: 30 seconds. 60 items. Maximum score 60. Exercise 2: 30 seconds. 48 items. Maximum score 48. Exercise 3: 30 seconds. 60 items. Maximum score 60. Exercise 4: 30 seconds. 60 items. Maximum score 60. Exercise 5: 30 seconds. 96 items. Maximum score 96. Exercise 6: 30 seconds. 48 items. Maximum Score 48. Exercise 7: 30 seconds. 48 items. Maximum score 48. Scored as number correct in time limit. Also scored total multiple responses, total omissions and total incorrect per sub-test.	Parallel forms for symbolic = .86 Parallel forms for non-symbolic = .86 Combined = .90
Mathematical Ability	Numerical Operations	Wechsler Individual Achievement Test (WIAT II) (Wechsler, 2005)	Day 1, Test 2	20 minutes. 26 items. Scored as number of items correct. Maximum score = 26 (n.b. items 1-6 were removed from the list from T1 and T2).	Cronbach's alpha = .88
	One Minute Addition	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.	Day 1, Test 4	One Minute. 60 items. Maximum score = 60.	Cronbach's alpha = .93
	One Minute Addition Extra		Day 1, Test 5	One Minute. 60 items. Maximum score = 60.	Cronbach's alpha = .91
	One Minute Subtraction		Day 1, Test 6	One Minute. 60 items. Maximum score = 60.	Cronbach's alpha = .88
	One Minute Subtraction Extra		Day 1, Test 7	One Minute. 60 items. Maximum score = 60.	Cronbach's alpha = .92
Fraction Understanding	Fraction Writing	Experimental	Day 1, Test 8	Untimed as read out by researcher. 10 items. Scored as number of fractions written correctly. Maximum score = 10.	Cronbach's alpha = .77
Mathematical Ability	One Minute Multiplication	Adapted from Westwood, P., Harris-Hughes,	Day 2, Test 1	56 Items. One Minute. Maximum score = 56.	Cronbach's alpha = .93

	One Minute Division	M., Lucas, G., Nolan, J., & Scrymgeour, K. (1974).	Day 2, Test 2	56 items. One Minute. Maximum score = 56.	Cronbach's alpha = .95
Mathematical Ability	Maths Reasoning	Wechsler Individual Achievement Test (WIAT II)	Day 2, Test 3	Untimed, led by researcher, task took approximately 15 minutes. 19 items. Maximum score = 19.	Cronbach's alpha = .76
Ordinality	Ordinality dots	Experimental	Day 2, Test 4	90 seconds. 80 items. Maximum score = 80.	
	Ordinality digits	Experimental	Day 2, Test 5	90 seconds. 80 items. Maximum score = 80.	
Numerical Knowledge	Conversion (Money, time, length)	Adapted from Eggenberger Rechentest 3+ (Holzer et al., 2007)	Day 2, Test 6	20 seconds for the money task.45 seconds each for the time and length tasks. 5 items each. Maximum score = 15.	Cronbach's alpha = .87
Numerical Knowledge	Number line estimation (three-digit numbers, 0-1000)	Experimental	Day 2, Test 7a	Untimed task led by researcher (approximate time 5 minutes). 1x10 XXX numbers. Scored as difference as percentage of number line from the target number for each number by participant.	Cronbach's alpha = .79
Fraction Understanding	Number line estimation (fractions, 0-1)	Experimental	Day 2, Test 7b	Untimed task led by researcher (approximate time 5 minutes). 1x10 fractions. Scored as difference as percentage of number line from the target number for each number by participant.	Cronbach's alpha = .79

Table 4. Individual testing: Tests used in the test battery at Time 3

Construct	Test	Standardised	Test Number	Administration Time, Items, Maximum Score	Reliability
Numerical Knowledge	Number Matching	Experimental	Individual Test 1	Computerised Task. Administration time around 7-8 minutes, dependent on participant speed. 168 items. Reaction time and accuracy data recorded by trial. Maximum score = 168.	Bivariate Pearson's Correlation between means for odd and even trials for each participant r =.95, p <.001
Reading	Word Reading and Pseudo-word reading	Test of Word Reading Efficiency (TOWRE-2)	Individual Test 2	45-seconds per test. Scored as total number of words read correctly in 45 seconds. Pseudo Word Reading Maximum score = 63.	Test-retest reliability for the Sight Word Efficiency task in children aged 8 on Form B was .92 and for the Phonemic Decoding Efficiency task in children aged 8 on Form B was .93 (Wagner et al., 2011)

Mathematical Conceptual Understanding	Mathematical Conceptual Understanding	Experimental	Individual Test 3	Computerised task. 24 items. Time taken around 8 minutes, dependent on participant speed. Reaction time and accuracy recorded by trial. 8 items required explanation. Maximum score correct = 24	Bivariate Pearson's Correlation between means for RTs on odd and even trials for each participant returned an r =. 41, p <.001. Cronbach's alpha (number of items correct) =. 79
Numerical Knowledge	Single Digit Matching	Experimental	Individual Test 4	Computerised task. Time taken around five minutes. Reaction time and accuracy recorded by trial. Maximum score correct = 144.	Bivariate Pearson's Correlation between means for odd and even trials for each participant returned an r = .96, p <.001

2.1 Group testing

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. Each task demonstration was shown prior to task completion.

2.1.1 Magnitude comparison

These tasks were administered as a part of the group testing during Day 1 and consisted of symbolic and non-symbolic subtests. A total of three digit comparison (symbolic) and six dot comparison (non-symbolic) subtests were administered. Symbolic and non-symbolic items were presented across two A5 booklets with six rows of items presented on each page and one pair in 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 item was given prior to each subtest.

Booklet one contained two symbolic comparison subtests (including one subtest as practice) and two non-symbolic subtests (including one as practice). The first subtest was a practice subtest comparing digits (symbolic), preceded by two worked through examples. The second exercise was a practice subtest comparing dots (non-symbolic comparison) and was also preceded by two worked examples. The two practice subtests contained 48 items each. This was followed by exercise 'Number 1' (symbolic) and exercise 'Number 2' (non-symbolic), both preceded by two worked examples. The 'Star Exercise' from Time 2 was not repeated at Time 3.

Booklet two contained four dot and one digit comparison subtests. It started with 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 each. More information on the items (average and SD of problem size, distance, number range and ratio) for these tasks can be found in Appendix 1. The tests were taken from Göbel, Watson, Lervåg and Hulme (2014). The booklets were designed so that the children could not see the first 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 those pages had no assessment data on).

Symbolic comparison tasks. 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 of one to four, or 'far' 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 Practice 1, 48 items were presented. The pairs of these number distances were mixed (e.g. included far and close distances), with an average problem size of 9.71, and a range of 14. In the two proceeding tasks (E1 & E4) 60 items were presented, one subtest presented number pairs with far distance (E1) and the other subtest number pairs with close distance (E4), both with an average size of 10 (range=5).

Non-symbolic comparison task. 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. The display size varied from five to 13 dots. For three of the non-symbolic comparison tasks there were 48 items (E2, E6, E7) and the pairs of dots 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). E3 had 60 items and a close distance and E5 had 96 items (increased from 60 at Time 1) and a far distance. Both had a problem size of 15 and all dots in E3 and E5 were of the same size.

2.1.2 Arithmetic

Numerical operations. This test contained 26 questions from the Wechsler Individual Achievement Test 2nd Edition (Wechsler, 2005), adapted for group use. Items 1-6 from T1 and T2 testing were removed. (Therefore test

item 1 at T3 had been test item 7 at T1 and T2, etc.) The presentation of the remaining items from T2 remained unchanged for consistency. Nine new items were added at Time 3. Items were written in Century Gothic font, font size 16. Two new items of whole number subtraction were added and presented as column subtraction, one item of whole number addition, two items of whole number division, one of whole number multiplication, one fraction item and two decimal items were also added. Items 1-9 were included on page 1, 10-17 page 2 and 18-26 on page 3. For more information about items by type of operation please see Appendix 2.

One-minute addition. This was set out over three A4 pages. The first page featured an example (1 + 1 =). This was followed by 60 items set out over two pages in two columns, typed in Calibri (body) font size 24. All questions were single-digit (one to nine) presented in a written format e.g. 2 + 1 =. The questions began with easier items (e.g. 2 + 1 =) and gradually increased in difficulty (e.g. 7 + 6 =). Half of the items included carrying on the addition. This test remained the same from T2. The questions were displayed in two columns, going down the page. The questions were set out horizontally (not columnar).

One-minute addition extra. This was set out over three A4 pages. The first page featured an example (10 + 7 =). This was followed by 60 items set out over two pages in two columns, typed in Calibri (body) font size 24. All questions were double-digit (1 to 96) plus single-digit (1 to 8) questions presented in a written format e.g. 12 + 2 =. The final 30 (of the 60) items were added at Time 3 (all on page 3). None of the items required carrying over to the decade. The questions were displayed in two columns, going down the page. The questions were set out horizontally (not columnar).

One-minute subtraction. This was set out over three A4 pages. The first page featured an example (3 – 1=). This was followed by 60 items set out over two pages 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 =. The questions began with easier items (e.g. 2 – 1 =) and gradually increased in difficulty (e.g. 7 – 6 =). 24 items of the 60 included borrowing on the subtraction questions. This test remained the same from T2. The questions were displayed in two columns, going down the page. The questions were set out horizontally (not columnar).

One-minute subtraction extra. This was set out over three A4 pages. The first page featured an example (19 – 2 =). This was followed by 60 items set out over two pages in two columns, typed in Calibri (body) font size 24. All

questions were double-digit (11 to 98) minus single-digit questions presented in a written format e.g. 71 – 1 =. None of the items required 'borrowing' from the decades. 30 (of the 60) items were added at T3 (all on page 3). The questions were displayed in two columns, going down the page. The questions were set out horizontally (not columnar).

One-minute multiplication. This was set out over three A4 pages. The first page featured an example (3 x 5 =). This was followed by 56 items set out over two pages 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. 5 x 2 =. The questions began with easier questions (e.g. 9 x 2 =) and gradually increased in difficulty (e.g. 7 x 8 =). This was a new test introduced at T3. The questions were displayed in two columns, going down the page. The questions were set out horizontally (not columnar).

One-minute division. This was set out over three A4 pages. The first page featured an example (25 ÷ 5 =). This was followed by 56 items set out over two pages in two columns, typed in Calibri (body) font size 24. All questions were double-digit divided by single digit questions presented in a written format e.g. 10 ÷ 2 =. The questions began with easier questions (e.g. 8 ÷ 2 =) and gradually increased in difficulty (e.g. 72 ÷ 8 =). This was a new test introduced at T3. The questions were displayed in two columns, going down the page. The questions were set out horizontally (not columnar).

Mathematical reasoning. This test was taken and adapted from the WIAT II mathematical reasoning subtest. We adapted it to group use by only selecting a subset of items (19) and by giving children an answer sheet. This answer sheet was set out in three columns: column one (1cm wide, length varied by question graphic) contained the question number, column two (9cm wide, length varied by question graphic) contained the graphic related to the question taken directly from the WIAT II (Wechsler, 2005). All graphics were presented in monochrome. The final column (8.4cm wide, length varied by question graphic) was blank for answers, except for four items where an answer prompt was included, as it was in the WIAT II test. The questions are included in Appendix 3.

2.1.3 Ordinality

In both ordinality tasks, there were 80 items. Each item consisted of three sets of numerosities between 1 and 9, which were either in ascending order (e.g., 2-4-6) or not in order (e.g., 2-6-4). Ascending triplets contained the

same combination of numerosities in both tasks. On each page there were 10 items set out in two columns with each column containing five items. 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 numerosities 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.

Dot Ordinality. The dot ordinality task was administered as part of the group testing during Session two. The dots were presented in boxes 2.86cm by 2.22cm. Each item consisted of three boxes with dots. Each box contained 1-9 dots. To ensure that magnitude was more salient than the physical features of the stimuli throughout the task, the overall surface area of the dots was either correlated or anti-correlated with the number of dots (i.e. surface area either increased or decreased with the number of dots), thus dot size varied between boxes. The nonsymbolic ordinality task included 41 ascending triplets, their number ranged between 3 and 6 on each page (see Appendix 4).

Digit Ordinality. This task was administered directly after the dot ordinality task. It also consisted of 80 items, 10 items per page (two columns of five items). Each item consisted of three Arabic digits (Arial, font size 48) each of which were presented in a box 2.86cm by 2.22cm. The symbolic ordinality task included 35 ascending triplets, their number varied between 2 and 6 on each page (see Appendix 5).

2.1.4 Numerical knowledge

Conversion task. This was presented as the final subtest on Day 2 in Booklet 1. Money, Time and Length conversion were examined, using items adapted from the Eggenberger Rechentest (ERT 3+) (Holzer, Schaupp & Lenart, 2007). The first part consisted of money conversion items which were presented on one page. The next page featured time conversion items and the last page length conversion items. There were five items for each subtest in Comic Sans, font size 20pt. Please see Appendix 6 for a list of items.

Number line estimation (0-1000). An A4 landscape booklet containing 22 pages was used. The first 11 consisted of the whole-number number line estimation task. The next 11 pages were used for the fraction line estimation – see below. The first page had a space for the child's name in the top right-hand corner, the title "Number Line Task" and an explanation of the task in Calibri 18 font. The practice number 500 was central on the

page in Calibri 36 font. The number line measured 261mm in length. The digit '0' was on the left-hand side of the line and the digits '1000' on the right, as shown below in Figure 1.

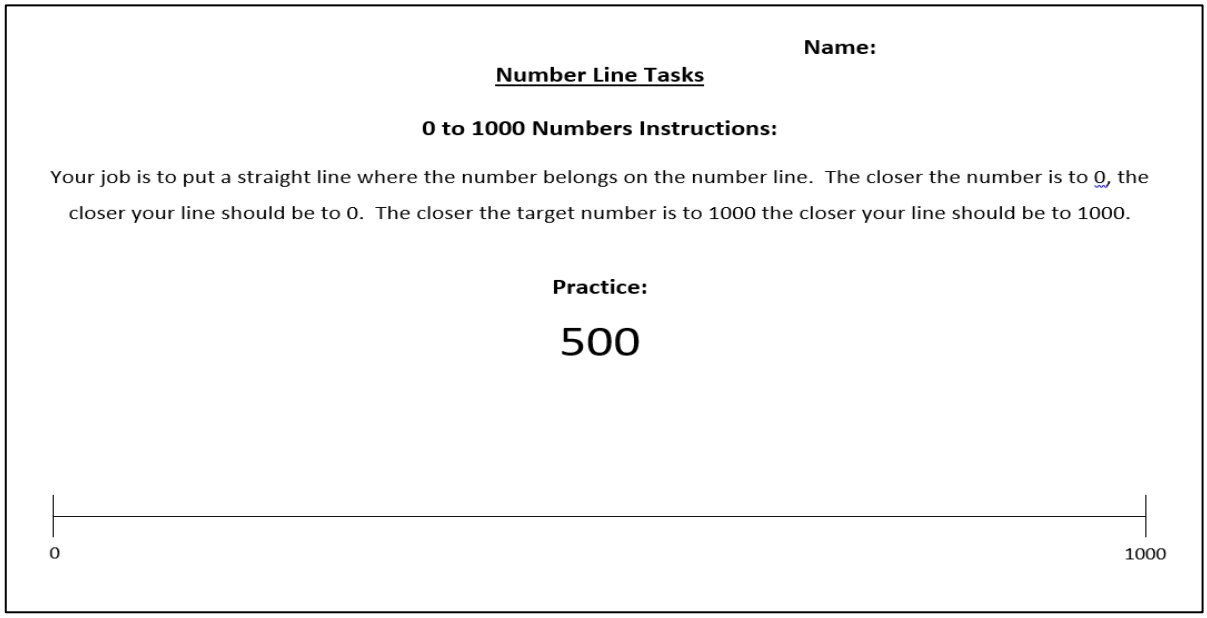

Figure 1: Number Line Task (0-1000), practice item

Each new number was presented on a new single sided A4 page, with the number and the number line in the same place on each page, as shown below in Figure 2.

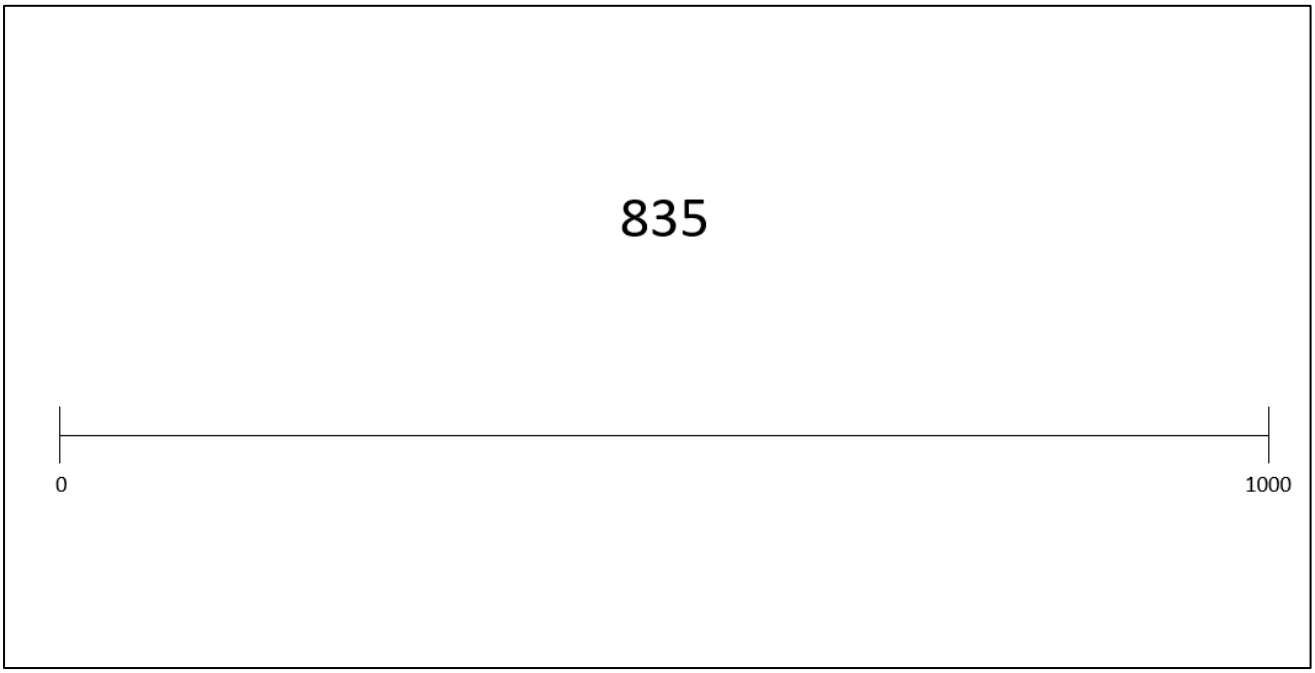

Figure 2: Number Line Task (0-1000), first test item

The target numbers were 835, 103, 721, 52, 475, 962, 508, 346, 613 and 297.

2.1.5 Fraction understanding

Fraction writing. This was a new task at T3. This task consisted of 10 fraction numbers in the form of Arabic digits. The format was the same as used for whole number writing at T1 and T2. The task was presented with a box for number entry (1.8c x 12 cm) next to a familiar cartoon illustration, for example, a carrot, a mouse, a tick, a chair, a dog etc, in order to help children with place keeping. Illustrations were displayed in a 1.8x4cm box on the left-hand side of the page. The full list of fraction writing items in the order they were presented, together with the accompanying descriptor pictures, and details of the test are attached in Appendix 7.

Fraction number line estimation (0-1). The last 11 pages of the number line estimation booklet (see above) contained the fraction number line estimation task. The first page had the title "Fraction Instructions" and an explanation of the task in Calibri 18 font, as shown in Figure 3. The number line was 261mm in length. The digit '0' was on the left-hand side of the line and the digit '1' on the right. There was no practice item. Each new number was presented on a new single sided A4 page, with the number and the number line in the same place on each page. The target fractions were:

$\frac{1}{4}$ $\frac{4}{6}$ $\frac{2}{5}$ $\frac{1}{2}$ $\frac{2}{8}$ $\frac{4}{10}$ $\frac{3}{4}$ $\frac{2}{3}$ $\frac{6}{8}$ $\frac{2}{4}$

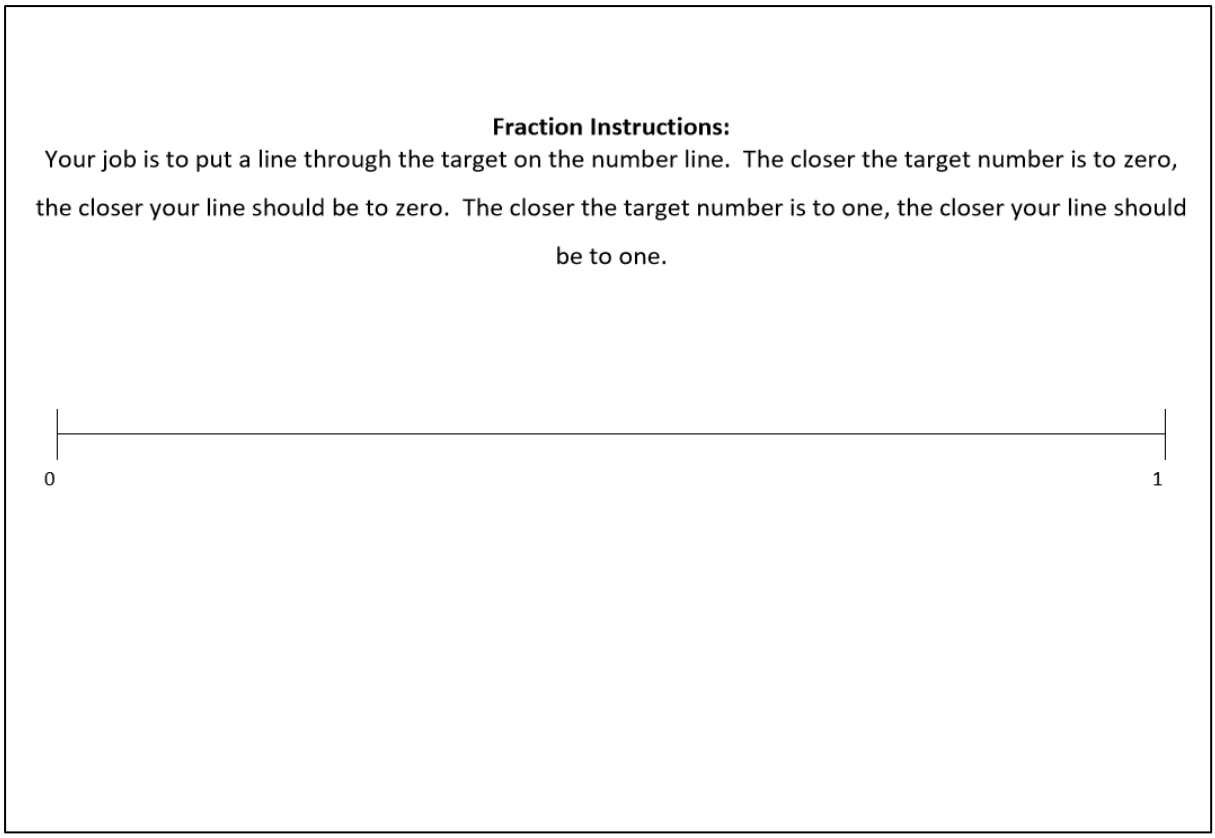

Figure 3: Presentation of Fraction Number Line Task

2.2 Individual testing

Additional materials for individual testing

Record Booklet. A record booklet containing individual record forms for each of the standardised tests and non-standardised tests and an overall cover sheet was used. The cover sheet contained the child's date of birth, gender, handedness, testing dates, general observations and individual test scores. This was presented as an A4 document in portrait orientation. Attached to the form were the participant's conceptual understanding form and TOWRE record forms. The Conceptual Understanding Record Sheet was an A4 portrait orientated sheet detailing the child's explanation for their selection of whether the preceding item helped solve the second item. An example of the sheet is included in Appendix 8A. A TOWRE record sheet (for TOWRE form B) was included for each participant and was attached to the record form.

Computer. A 15.6-inch Dell Inspiron 15 laptop (resolution 1920 x 1080 pixels) and keyboard were used for the number matching, conceptual understanding and single-digit matching tasks.

Headphones. Sony headphones (Sony MDRZX310) were used for the Number Matching and Single Digit Matching Tasks.

Additional Keyboard. An additional keyboard was used and attached to the computer for child use during individual testing.

Stopwatch. A stopwatch was used for administering the speeded reading tasks. The stopwatches were ATM Model 136.

2.2.1 Number matching

The task was presented in PsychoPy v1.85.3 (Peirce, 2007). Numerical stimuli were made out of a combination of the digits 1 to 9, with the omission of 'seven' because it is disyllabic. Visual stimuli were presented in black on a white background in Arial font with a proportional height of 0.3, compared to overall screen size. Auditory stimuli were recorded by a female native speaker. All numbers were trimmed to remove excess time before and after the spoken number. The average duration per number word was 1.2 seconds.

Thirty different auditory targets (no decade numbers, no ties) were presented. Teen targets were presented four times (target presented twice, an inversion and a lexical condition), double digit and triple digit numbers were presented six times (target presented three times, and three error conditions).

Following six practice trials, there were 168 trials (48 new items were added at T3) consisting of 24 teen items (12 matching and 12 distractors,) 96 two-digit number items (48 matching and 48 distractors) and 48 three-digit items (24 matching and 24 distractors). The order of experimental trials was pseudo-randomized with the restriction that identical number words were never presented consecutively and no more than three trials with the same expected response appeared in immediate succession.

To avoid a bias toward "no" responses, on 84 trials the verbal number words were followed by the matching Arabic number. The remaining 84 trials comprised six possible non-matching distractors:

(1) the decade matched with the target whereas a different digit appeared in the unit position D+U–, e.g., *twenty-three* → *25*, occurred 8 times.

(2) the unit digit of the target appeared at the decade position, whereas the unit position was incorrect, D–U+, e.g., *twenty-three* → *35*, occurred 8 times.

(3) an inverted distractor D+U+ e.g., *twenty-three* → *32*, appeared 22 times.

(4) the units matched with the target whereas a different digit appeared in the decade position D–U+, e.g., *sixty-three* → *23*, occurred 8 times.

(5) a nonrelated distractor D–U–, e.g., *twenty-three* → *46*. occurred 22 times.

(6) U_D_Corr, errors on the units and decades with the hundreds digit correct, e.g. *nine hundred and thirty-two* →*936* occurred 8 times.

Each trial began with a blank screen displayed for 400ms, followed by the presentation of the auditory number. Immediately after the offset of the auditory stimulus, an Arabic number appeared on the screen. The Arabic number was displayed until the participant responded, with a maximum duration of 4 seconds. Participants were instructed to press the green-stickered button ("L" key on the right part of the keyboard) when the auditory and visual numbers matched and the red-stickered button ("A" key on the left part of the keyboard) in case of a mismatch.

2.2.2 Conceptual understanding

The task was presented in PsychoPy v1.85.3 (Peirce, 2007). The screen background was black with white numbers and white text presented in Arial font. The text size was proportional (set at 0.15) to screen size. The resolution was set at 1920 x 1080 pixels. The first line contained a sum with its answer followed by the text 'If you know that can it help you solve' followed by presentation of a sum without its answer. Participants pressed the green-stickered button ("L" key on the right side of the keyboard) for a 'yes' response and the red-stickered button ("A" key on the left side of the keyboard) for a 'no' response. On eight trials a further probe appeared on the screen 'can you explain why or not?' (details below). The next trial was initiated by the experimenter (by pressing the enter key).

There were 29 trials in total including five practice trials (one for each type of relationship: commutativity, identical, inverse, unrelated and sub-comp) and 24 experimental trials (with the following breakdown: 6 identical trials, 6 inverse trials, 6 unrelated trials, 2 sub-comp trial, and 4 commutativity trials). The list of trials can be found in Appendix 8B.

Explanations were probed eight times per participant, on items 6, 2, 5, 4, 8, 1, 3 and 7, explanations were recorded using a response sheet (see Appendix 8A).

2.2.3 Single-digit matching

This was a computer-based task with 144 trials and presented in PsychoPy v1.85.3 (Peirce, 2007). Numerical stimuli constituted of the digits 1, 2, 3, 6, 8 and 9. Visual stimuli were presented centrally in white on a black background in Arial font (height subtending 3 degrees of visual angle). Auditory stimuli were recorded by a female native speaker. All numbers were trimmed to remove excess time before and after the spoken number and length was adjusted to approximately 500ms for each number word.

There were three conditions:

1. Simultaneous presentation: visual number and auditory number presented at the same time (48 trials, SOA 0)
2. VA: Visual number presented 500ms before the verbal number (48 trials, SOA 500ms)
3. AV: Verbal number presented 500ms before the visual number (48 trials, SOA -500ms)

The 48 trials within each condition were divided into 24 trials where the presented numbers were matching (e.g. 1 and 'one') and 24 where they were non-matching (e.g. 1 and 'three'). 12 of the non-matching number pairs had a 'far' distance and 12 had a 'near' distance. The full list of items is included in Appendix 9. The task was always presented in the same order. Three breaks were included with a 'take a break' screen after 36, 72 and 108 trials. The task included 12 practice items. On each trial, a fixation cross (white, 48 pt) was displayed first in the centre of the screen. After 500ms the fixation cross disappeared and the first stimulus was presented for 500ms (AV: spoken number word, VA: Arabic single digit at the centre of the screen, simultaneous presentation: spoken number word and Arabic single digit at the centre of the screen), then in AV and VA conditions the second stimulus (AV: Arabic single digit, VA: spoken number word) was presented for 500ms directly after the offset of stimulus 1. The next trial started automatically after a response or if a response was not given within 4 seconds after the end of the presentation of the second stimulus. Children were instructed to press the 'L' button (with the green sticker) if the numbers were matching and the 'A' button (with the red sticker) if the numbers were non-matching.

.

2.2.4 Word reading

Both the Sight Word Efficiency (SWE) and Phonemic Decoding Efficiency (PDE) subtests of the Test of Word Reading Efficiency (Wagner et al., 2011) were used. For both tasks, the practice trials included a single column of

eight items. The word reading task was a page of four columns with 27 items in each with the items increasing in length and difficulty (108 items in total). The non-word reading task consisted of three columns with 22 items in each (66 items in total).

3. Procedure

All children took part in two one-hour group testing sessions, one session 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 always took place prior to any individual testing. Some children were 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 four tests and group testing sessions comprised of fifteen tests divided across five Booklets (see Tables 3 and 4).

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 in all timed tasks (when a green traffic light was displayed on the screen before they were allowed to start any of the group tasks) and to stop and put their hands up as soon as the researcher said stop. 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. 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: Numerical Operations, one-minute addition, one-minute addition extra, one-minute subtraction, one-minute subtraction extra, fraction writing

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

Day 1 Booklets 1 and 3

Symbolic and non-symbolic magnitude comparison

Participants were given A5 booklets (two separate booklets were given out during day one of group testing) containing a total of nine magnitude comparison tasks (including 2 practice 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 in those times were crossed out by the researcher. This happened rarely and the extra completed item was disregarded when the data was entered. 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. In the magnitude comparison tasks 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 scores for each subtest were as follows: 48 points for Practice 1 and 2; 60 points for Exercise 1, 48 points for Exercise 2, 60 points for Exercises 3 and 4, 96 points for Exercise 5, and 48 points each for Exercise 6 and 7.

Day 1, Booklet 2

Children were provided with an A4 booklet containing the numerical operations task, taken from Wechsler Individual Achievement Test 2nd Edition (WIAT II), and the one-minute addition, one-minute addition extra, one-minute subtraction, one-minute subtraction extra and fraction writing tests.

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. Children were given up to 20 minutes to complete as many items as they could. 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. When time was up, 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 26.

One-minute addition and subtraction

One-Minute Addition, One Minute Addition Extra, One Minute Subtraction, One Minute Subtraction Extra

These tasks comprised of 60 addition items in the one-minute addition sub-test, 60 addition items in the addition extra sub-test, 60 subtraction items in the one-minute subtraction sub-test and 60 subtraction items in the subtraction extra sub test. Children were given 60 seconds for each of these four sub-tests to complete as many items in that sub-test as they could.

Prior to task completion children were shown an example on the PowerPoint presentation which corresponded to the practice example in their booklets. 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. Then the children were given another 60 seconds to answer some more difficult addition questions (preceded by an example completed together by the entire class). This process was repeated for one-minute subtraction and one-minute extra subtraction (questions included some double-digit minuends with all single-digit subtrahends).

Coding and scoring. One point was given for each correct item. The total possible correct score was 60 for addition, 60 for subtraction, 60 for addition extra and 60 for subtraction extra. Each of the four sub-tests was scored individually.

Fraction writing

Children were instructed to write the Arabic forms of dictated numbers next to a given illustration. For example, the researcher would say “write one half next to the chair”. Each number was spoken once but could be repeated a further two times, on request. This task was not timed. The fractions were read as detailed in Appendix 10.

Coding and scoring. Children were given one point for each number written correctly. The maximum possible score was 10.

3.2 Day 2 Testing

Booklet order:

Day 2, Booklet 1: one-minute multiplication, one-minute division, mathematical reasoning, dot ordinality and digit ordinality, conversion (money, time, length)

Day 2, Booklet 2: Number line estimation 0-1000, number line estimation 0-1

Day 2, Booklet 1

One-minute multiplication and division

For these task children were asked to complete as many of the 56 multiplication questions or 56 division questions as they could within the given time limit of 60 seconds per task. Children were instructed to go down the columns on the page. 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. Prior to task completion children were shown an example on the PowerPoint presentation which corresponded to the practice example in their booklets. This example was completed as a group.

Coding and scoring. One point was given for each correct item. The total number of correct items was calculated for the one-minute multiplication and one-minute division separately. Total maximum score for each of these tests was 56.

Mathematical reasoning

Children were told that the researcher would read some questions out aloud and that the space around the answers or at the back of the booklet could be used to do working out. They were told not to rush ahead and wait until each question was read out before answering. Researchers gave approximately one minute for each item, or until all children had completed it (whichever was sooner). For items and questions see Appendix 3.

Coding and Scoring. One mark was given for each item. The maximum available score was 19 (t3mr_old). Later we decided to exclude item 11 and calculate a new total score correct for each child (t3mr; max = 18).

Ordinality (dots and digits)

Dot Ordinality. Children were asked to tick the row of three sets of dots if the dots were increasing in amount, and to draw a line through the row if the dots were not ordered by increasing amount. 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 practice 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 amount. The possible maximum score for this task was 80.

Digit Ordinality. 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 (left column first, then move on to the right column). 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.

Conversion

The task started with monetary conversion. Children were asked to try to convert different amounts of money from pounds to pence and pence to pounds. Children were given twenty seconds to answer as many of the money items as they could. The second task was time conversions. Children were given forty-five seconds to answer as many of the time conversion items as possible. The final subtest was length conversion. Children were given forty-five seconds to answer as many of these items as possible.

Coding and scoring. One point was given for each correct item. The possible maximum score for each subtest was 5 and for the overall total correct score 15.

Day 2, Booklet 2

Number line estimation (0-1000; 0-1)

Children were presented with the A4 landscape number line booklet containing the whole-number number line task and the fraction number line task.

Number line 0-1000. Children were told to put a vertical line through the printed number line where they thought the number should be. They were told that the closer the number was to zero, the closer on the line their number should be to zero and the closer the number was to thousand the closer to that end of the number line their vertical line should be. Children were instructed to make one vertical line straight through where they thought the number should be. The number 500 was done as a practice trial. The experimenter checked after the practice trial that children had used vertical lines and only one line. After the practice trial the children were instructed to put a vertical line through the number line to indicate the location on the number line for each of the following numbers and to move through the items at their own pace. They were instructed to stop after they had placed 297 (on the last whole number page) and to put their hands on their head when they had finished.

Fraction number line 0-1. Children were told to put a mark on the line where they thought the number should be. They were told that the closer the number was to zero the closer to zero on the number line their marking should be to zero and the closer the number was to 1 the closer their marking on the number line should be to 1. Children were instructed to make one vertical line straight through the middle where they thought the number should be. There were no practice trials because the children had already completed the whole number line task. The numbers were read out once by the experimenter but could be repeated up to three times on request from the

children. The experimenter waited until all the children had completed each item before moving on to the next page. Children were instructed to wait for the instruction of the experimenter to turn over before going on to the next page.

Coding and scoring for both number lines tasks. Separately for the two tasks, for each target number we calculated where on the line the correct response would be (in mm from the left endpoint) assuming a linear left-to-right placement. We then converted this into a percentage of the whole line length (261mm). These two values are constant across participants. For each child and each target number we then measured the distance in mm from the left endpoint of the line to the point where the vertical line made by the child crossed the line (in data dictionary: t3nltr). We converted this also into a percentage of the whole line length (t3nlpp). We the calculated the percentage error per target number per child by subtraction the correct percentage of the line length from the percentage placement for the child (t3nlte). For example, for the number 835, the correct placing was 217mm (84% of the number line), if the participant's vertical line crossed the number line at 250mm, that would be 96.15% of the number line. The percentage error would thus be +12.65%. This percentage error can be positive or negative. Thus we also converted the percentage error into an absolute percentage error (t3nlce). Accuracy (t3nlpa) was calculated as 100 - the absolute percentage error on each number.

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. The researcher was seated to the right of the child. The computer and additional keyboard were placed directly in front of the child. Children were 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. Details of the child's gender and the dates of testing were also collected. Auditory stimuli were conveyed bilaterally through headphones.

Test order*:*

Number matching, TOWRE reading (words and pseudowords), conceptual understanding, single-digit matching

Number matching

This was a computer-based task. The task was run in Psychopy v1.85.3 (Peirce, 2007) which was saved to the desktop of the laptops used before the visits took place. All participants received oral instructions from the experimenter. The researcher explained the goal was to be as fast and accurate as possible. The child sat in front of the laptop and heard a series of numbers through the headphones while being presented with numbers on the screen. Volume was set consistently to 20 and only adjusted at the child's request. The child's task was to decide on each trial whether the spoken number word and the visually presented Arabic digit string matched. Children were instructed to press the right ('L') key (which had a green sticker on it) when the items matched and to press the left ('A') key (with a red sticker on it) when the items did not match. Six practice trials with feedback were included at the beginning of the task. If the child had not understood the game it was paused and re-explained. This happened in very few (less than ten) cases.

Coding and scoring. The child's responses (reaction time and accuracy) were recorded automatically for each trial. The maximum number of trials correct was 168.

Reading: Sight word efficiency

Sight Word Efficiency (TOWRE-2, Form B) (Wagner et al., 2011)

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 stimulus 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

Phonemic decoding (TOWRE-2, Form B) (Wagner et al., 2011)

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.

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

Conceptual understanding

The instructions were read out to the child and they were asked to press the green stickered key ('L' key) when the sum at the top of the screen (presented with its answer) helped solve the lower sum (presented with no answer) and to press the red stickered key ('A') key when it did not help. The task took around 5 minutes to administer. On eight trials, the children were asked to explain their choice and this was recorded by the researcher on the form contained in Appendix 8A.

Coding and Scoring: We recorded reaction time and accuracy as well as their explanations. The maximum score for correct explanation was eight and for overall items correct the maximum was 24.

Single-digit matching

The instructions were read out to the child and were to press the green stickered key ('L' key) when the auditory and visual number were the same and to press the red stickered key ('A') key when the auditory and visual number were different. The task took around 6 minutes to administer. At the three breaks in the task, children were instructed to stretch and wiggle their hands and look away from the screen.

Coding and Scoring: The child's responses were recorded automatically by reaction time and items correct. The maximum score for this subtest was 144.

4. Reference list

Holzer, N., Schaupp, H., & Lenart, F. (2007). *Eggenberger Rechentest (ERT 3+)*. Hogrefe.

National Statistics. (2019). *School data comparison*. compare-schools-performance service.gov.uk

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Wagner, R., Torgensen, J. K., & Rashotte, C. (2011). *Test of Word Reading Efficiency—Second Edition (TOWRE-2)*. Pearson.https://www.pearsonclinical.co.uk/Psychology/ChildCognitionNeuropsychologyandLanguage/ChildLanguage/TOWRE2/TestofWordReadingEfficiencySecondEdition.aspx

Wechsler, D. (2005). *Wechsler Individual Achievement Test—Second UK Edition (WIAT-II UK)*. Pearson.

Appendix 1: 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	96	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	9.02	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 2: Numerical operation

Mathematical Operation	Items	Page	New item for Time 3?
Addition (9 items)	3+3=, 8+5= 2+3+1+4= 41+14= 37+54=	1	
	698+426= 57 + 32 + 94 + 48=	2	
	753+219= 0.2+0.8=	3	yes yes
Subtraction (9 items)	4 – 2= 10 – 6= 68 – 43=	1	
	120 – 15= 80 – 56=	2	
	978-532= 705-489= 5.47-2.31= $\frac{7}{8} - \frac{3}{8} =$	3	yes yes yes yes
Multiplication (4 items)	8 x 5 =	1	
	7 x 6 = 24 x 5=	2	
	297 x 7=	3	yes
Division (4 items)	16÷2= 69÷3=	2	
	800÷4= 744÷6=	3	yes yes

Appendix 3: WIAT mathematical reasoning

WIAT Item #	T3 Item #	Original WIAT question / adapted question used at T3
Age 6		
9		When you are counting, which number comes next after ten? [Question removed]
10	1	Neil had five marbles. Then his mother gave him three more marbles. How many marbles did he have then?
11		If Angie has one bowl of food for each dog, how many dogs will not have a bowl? If each bowl is given to a dog, how many dogs will not have a bowl? [Question removed]
12		Each small square equals one square unit. How many square units are shaded? [Question removed]
13		If two of these ducks flew away, how many would be left? [Question removed]
14		If you are counting in order, which of these numbers would you say first? [Question removed]
15	2	Point to the second apple from the bowl. Circle the second apple from the bowl.
Age 7		
16	3	When you are counting, which of these numbers do you say first?
17		Yvonne used beads to make a pattern on the pegs. Part of her pattern looked like this. How many beads should Yvonne put on the empty peg to continue the pattern? [Question removed]
18		Marcus used beads to form a pattern on the pegs. Part of his pattern looked like this. How many beads should Marcus put on the empty peg to continue the pattern? [Question removed]
19	4	How long is the pencil? [Pencil image swapped to cm side of ruler]
20		How many pence does it take to equal the value of one pound? [Pound image updated] [Question removed]
Age 8		
21		What time is shown on this clock? [Question removed]
22	5	On what day of the week is the 14th? [Calendar image changed so week begins on Monday instead of Sunday]
Age 9		
23	6	This graph shows the number of books given to the city library by students from four different schools during the National Book Week. Going across the bottom of the graph, the schools are (point as you say it) Central, Johnson, West, and Eastman. The numbers on the side show the number of books given by each school. How many books did Eastman school give?
24		Which school came third in the number of books given? [Question removed]
25	7	Five ducks were swimming in a pond. Three flew away, then two more came to swim. Then how many ducks were in the pond?
26	8	What number goes in the empty circle?
27		Which is worth more: seven pence, six five p's (5p) or a ten p (10p)? [Question removed]

Age 10		
28		These blocks are each divided into four squares. Circle the block where ¾ of the squares are shaded. These blocks are each divided into four squares. Circle the block where three quarters of the squares are shaded. [Question removed]
29	9	What time is shown on this clock?
Age 11		
30	10	Robert has six stones. Together Robert and Max have fifteen stones. How many stones does Max have?
31	11	How much money is this? [Currency images updated]
32	12	If you were counting in order, which number you would you say last? If you were counting in order, circle the number you would you say last?
33		If you tossed a coin ten times, how many times would the coin be likely to land on heads? [Question removed]
34	13	If today is the 3rd of the month, and John's cousin will come to visit on the 17th, how many weeks must John wait until his cousin arrives? [Calendar image changed so week begins on Monday instead of Sunday]
Age 12-13		
35	14	Erik had four pounds on Monday. On Tuesday he earned two pounds mowing the lawn. On Thursday he spent three pounds at the cinema. How much money did he have left?
36	15	What is the missing number?
37	16	Mrs Ryan's classroom has four rows of desks. Each row has the same number of desks. There are a total of twenty-four desks. How many desks are in each row?
Age 14-21		
42	17	Put these fractions in order from smallest to largest.
48	18	What is the next number in this pattern?
53	19	Jan went to sleep at 10:30 p.m. and woke up at 7:00 a.m. the next morning. How long did Jan sleep?

Appendix 4: Dot ordinality items

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

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

Appendix 5: Digit ordinality items

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

Appendix 6: Conversion items

Money Items

1. £ 1 = _____ p
2. £ 3 = _____ p
3. 800 p = £ _____
4. 670 p = £ _____ _____ p
5. £ 7 and 5 p = _____ p

Time Items

1. 60 mins = _____ h
2. 90 mins = _____ h _____ mins
3. 3 h = _____ mins
4. 2 h 20 mins = _____ mins
5. 4 h 10 mins = _____ mins

Length Items

1. 3 m = _____ cm
2. 1 km = _____ m
3. 1 cm = _____ mm
4. 70 cm = _____ mm
5. 3700 m = _____ km _____ m

Appendix 7: Fraction writing

	$\frac{1}{2}$
	$\frac{1}{4}$
	1/7
	1/3
	1/5
	$\frac{3}{4}$
	2/5
	7/10
	1 1/8
	2 5/6

Appendix 8: Conceptual understanding

A: Response sheet for explanations

Conceptual Knowledge Responses: Can you explain why or why not?					
Item	**Relationship**	**Answer**	**Correct explanation (or similar)**	**Incorrect explanation**	**"I don't know"**
63 - 31 = 32 63 - 31 =	identical	y	The numbers are the same		
23 + 24 = 47 32 + 24 =	Unrelated	n	The numbers are different		
63 - 31 = 32 63 - 32 =	sub comp	y	The answer has been swapped around		
23 + 24 = 47 23 + 24 =	identical	y	The numbers are the same		
63 - 31 = 32 63 - 13 =	Unrelated	n	The numbers are different		
23 + 24 = 47 24 + 23 =	Commutativity	y	The answer is the same when the numbers are in different order		
23 + 24 = 47 47 - 23 =	Inverse	y	It's the opposite sum		
63 - 31 = 32 31 + 32 =	inverse	y	It's the opposite sum		
Total correct responses					

B: Conceptual understanding items

Problem Number	Top Problem	Bottom Problem	practice	related	relationship
1 (practice)	1 + 7 = 8	7 + 1 =	y	y	commutativity
2 (practice)	8 - 2 = 6	8 - 6 =	y	y	sub comp
3 (practice)	3 + 4 = 7	2 + 6 =	y	n	unrelated
4 (practice)	25 - 12 = 13	25 - 12 =	y	y	identical
5 (practice)	23 + 16 = 39	39 - 23 =	y	y	inverse
1	23 + 24 = 47	24 + 23 =	n	y	commutativity
2	23 + 24 = 47	32 + 24 =	n	n	unrelated
3	23 + 24 = 47	47 - 23 =	n	y	inverse
4	23 + 24 = 47	23 + 24 =	n	y	identical
5	63 - 31 = 32	63 - 32 =	n	y	sub comp
6	63 - 31 = 32	63 - 31 =	n	y	identical
7	63 - 31 = 32	31 + 32 =	n	y	inverse
8	63 - 31 = 32	63 - 13 =	n	n	unrelated
9	76 - 32 = 44	76 - 44 =	n	y	sub comp
10	76 - 32 = 44	67 - 32 =	n	n	unrelated
11	76 - 32 = 44	76 - 32 =	n	y	identical
12	76 - 32 = 44	32 + 44 =	n	y	inverse
13	41 + 27 = 68	68 - 27 =	n	y	inverse
14	41 + 27 = 68	14 + 72 =	n	n	unrelated
15	41 + 27 = 68	27 + 41 =	n	y	commutativity
16	41 + 27 = 68	41 + 27 =	n	y	identical
17	23 + 36 = 59	23 + 36 =	n	y	identical
18	23 + 36 = 59	32 + 36 =	n	n	unrelated
19	23 + 36 = 59	59 - 23 =	n	y	inverse
20	23 + 36 = 59	36 + 23 =	n	y	commutativity
21	38 + 23 = 61	23 + 38 =	n	y	commutativity
22	38 + 23 = 61	38 + 23 =	n	y	identical
23	38 + 23 = 61	61 - 23 =	n	y	inverse
24	38 + 23 = 61	38 + 32 =	n	n	unrelated

Appendix 9: Single-digit matching

index	Select Trial	SOA	S1	S2	Correct Answer	Number size	distance	distance condition
1	VA	500	2	3	a	SMALL	1	DIFF-CLOSE
2	VA	500	6	6	l	LARGE	0	SAME
3	AV	-500	1	1	l	SMALL	0	SAME
4	AV	-500	8	3	a	LARGE	5	DIFF-FAR
5	VA	500	9	9	l	LARGE	0	SAME
6	AV	-500	1	2	a	SMALL	1	DIFF-CLOSE
7	SIM	0	1	6	a	SMALL	5	DIFF-FAR
8	AV	-500	9	9	l	LARGE	0	SAME
9	AV	-500	6	6	l	LARGE	0	SAME
10	SIM	0	8	3	a	LARGE	5	DIFF-FAR
11	SIM	0	1	1	l	SMALL	0	SAME
12	AV	-500	1	6	a	SMALL	5	DIFF-FAR
13	AV	-500	8	9	a	LARGE	1	DIFF-CLOSE
14	VA	500	1	1	l	SMALL	0	SAME
15	SIM	0	6	8	a	LARGE	2	DIFF-CLOSE
16	SIM	0	3	3	l	SMALL	0	SAME
17	VA	500	6	1	a	LARGE	5	DIFF-FAR
18	SIM	0	8	8	l	LARGE	0	SAME
19	AV	-500	6	6	l	LARGE	0	SAME
20	AV	-500	1	8	a	SMALL	7	DIFF-FAR
21	AV	-500	2	2	l	SMALL	0	SAME
22	AV	-500	9	6	a	LARGE	3	DIFF-CLOSE
23	SIM	0	1	8	a	SMALL	7	DIFF-FAR
24	VA	500	6	6	l	LARGE	0	SAME
25	VA	500	2	2	l	SMALL	0	SAME
26	SIM	0	1	1	l	SMALL	0	SAME
27	SIM	0	6	9	a	LARGE	3	DIFF-CLOSE
28	SIM	0	3	3	l	SMALL	0	SAME
29	AV	-500	1	3	a	SMALL	2	DIFF-CLOSE

30	VA	500	6	2	a	LARGE	4	DIFF-FAR
31	VA	500	3	1	a	SMALL	2	DIFF-CLOSE
32	VA	500	9	9	l	LARGE	0	SAME
33	AV	-500	9	2	a	LARGE	7	DIFF-FAR
34	SIM	0	9	2	a	LARGE	7	DIFF-FAR
35	SIM	0	8	8	l	LARGE	0	SAME
36	AV	-500	9	9	l	LARGE	0	SAME
37	AV	-500	2	1	a	SMALL	1	DIFF-CLOSE
38	AV	-500	9	3	a	LARGE	6	DIFF-FAR
39	SIM	0	8	6	a	LARGE	2	DIFF-CLOSE
40	VA	500	8	1	a	LARGE	7	DIFF-FAR
41	SIM	0	9	9	l	LARGE	0	SAME
42	AV	-500	9	8	a	LARGE	1	DIFF-CLOSE
43	SIM	0	1	1	l	SMALL	0	SAME
44	AV	-500	2	2	l	SMALL	0	SAME
45	SIM	0	3	3	l	SMALL	0	SAME
46	VA	500	6	6	l	LARGE	0	SAME
47	AV	-500	2	6	a	SMALL	4	DIFF-FAR
48	VA	500	9	9	l	LARGE	0	SAME
49	VA	500	2	2	l	SMALL	0	SAME
50	AV	-500	9	9	l	LARGE	0	SAME
51	SIM	0	2	6	a	SMALL	4	DIFF-FAR
52	SIM	0	9	3	a	LARGE	6	DIFF-FAR
53	AV	-500	6	6	l	LARGE	0	SAME
54	VA	500	3	2	a	SMALL	1	DIFF-CLOSE
55	SIM	0	1	2	a	SMALL	1	DIFF-CLOSE
56	SIM	0	2	9	a	SMALL	7	DIFF-FAR
57	SIM	0	9	9	l	LARGE	0	SAME
58	AV	-500	6	6	l	LARGE	0	SAME
59	AV	-500	2	9	a	SMALL	7	DIFF-FAR
60	VA	500	2	2	l	SMALL	0	SAME
61	VA	500	6	6	l	LARGE	0	SAME
62	AV	-500	2	2	l	SMALL	0	SAME

63	VA	500	8	3	a	LARGE	5	DIFF-FAR
64	SIM	0	6	6	l	LARGE	0	SAME
65	AV	-500	2	3	a	SMALL	1	DIFF-CLOSE
66	SIM	0	1	1	l	SMALL	0	SAME
67	VA	500	6	8	a	LARGE	2	DIFF-CLOSE
68	SIM	0	8	9	a	LARGE	1	DIFF-CLOSE
69	VA	500	1	6	a	SMALL	5	DIFF-FAR
70	SIM	0	1	3	a	SMALL	2	DIFF-CLOSE
71	VA	500	8	8	l	LARGE	0	SAME
72	AV	-500	2	2	l	SMALL	0	SAME
73	AV	-500	3	1	a	SMALL	2	DIFF-CLOSE
74	AV	-500	8	8	l	LARGE	0	SAME
75	VA	500	6	9	a	LARGE	3	DIFF-CLOSE
76	VA	500	2	2	l	SMALL	0	SAME
77	SIM	0	6	6	l	LARGE	0	SAME
78	AV	-500	3	8	a	SMALL	5	DIFF-FAR
79	SIM	0	2	2	l	SMALL	0	SAME
80	VA	500	9	2	a	LARGE	7	DIFF-FAR
81	SIM	0	9	9	l	LARGE	0	SAME
82	SIM	0	3	8	a	SMALL	5	DIFF-FAR
83	SIM	0	9	6	a	LARGE	3	DIFF-CLOSE
84	VA	500	1	8	a	SMALL	7	DIFF-FAR
85	AV	-500	8	8	l	LARGE	0	SAME
86	SIM	0	3	9	a	SMALL	6	DIFF-FAR
87	AV	-500	3	3	l	SMALL	0	SAME
88	SIM	0	2	2	l	SMALL	0	SAME
89	VA	500	8	8	l	LARGE	0	SAME
90	AV	-500	3	2	a	SMALL	1	DIFF-CLOSE
91	VA	500	3	3	l	SMALL	0	SAME
92	SIM	0	9	9	l	LARGE	0	SAME
93	VA	500	9	3	a	LARGE	6	DIFF-FAR
94	SIM	0	6	6	l	LARGE	0	SAME
95	VA	500	2	6	a	SMALL	4	DIFF-FAR

96	SIM	0	2	1	a	SMALL	1	DIFF-CLOSE
97	SIM	0	9	8	a	LARGE	1	DIFF-CLOSE
98	VA	500	8	6	a	LARGE	2	DIFF-CLOSE
99	AV	-500	3	9	a	SMALL	6	DIFF-FAR
100	SIM	0	2	3	a	SMALL	1	DIFF-CLOSE
101	AV	-500	6	1	a	LARGE	5	DIFF-FAR
102	VA	500	8	9	a	LARGE	1	DIFF-CLOSE
103	SIM	0	6	6	l	LARGE	0	SAME
104	AV	-500	6	8	a	LARGE	2	DIFF-CLOSE
105	VA	500	1	2	a	SMALL	1	DIFF-CLOSE
106	VA	500	3	3	l	SMALL	0	SAME
107	AV	-500	8	8	l	LARGE	0	SAME
108	VA	500	8	8	l	LARGE	0	SAME
109	SIM	0	6	1	a	LARGE	5	DIFF-FAR
110	AV	-500	1	1	l	SMALL	0	SAME
111	SIM	0	2	2	l	SMALL	0	SAME
112	AV	-500	3	3	l	SMALL	0	SAME
113	VA	500	1	1	l	SMALL	0	SAME
114	VA	500	2	9	a	SMALL	7	DIFF-FAR
115	AV	-500	3	3	l	SMALL	0	SAME
116	SIM	0	6	2	a	LARGE	4	DIFF-FAR
117	SIM	0	3	1	a	SMALL	2	DIFF-CLOSE
118	VA	500	3	8	a	SMALL	5	DIFF-FAR
119	SIM	0	2	2	l	SMALL	0	SAME
120	SIM	0	8	8	l	LARGE	0	SAME
121	AV	-500	6	2	a	LARGE	4	DIFF-FAR
122	AV	-500	6	9	a	LARGE	3	DIFF-CLOSE
123	VA	500	8	8	l	LARGE	0	SAME
124	AV	-500	8	8	l	LARGE	0	SAME
125	VA	500	9	6	a	LARGE	3	DIFF-CLOSE
126	VA	500	3	3	l	SMALL	0	SAME
127	VA	500	1	1	l	SMALL	0	SAME
128	AV	-500	1	1	l	SMALL	0	SAME

129	VA	500	1	3	a	SMALL	2	DIFF-CLOSE
130	AV	-500	9	9	l	LARGE	0	SAME
131	SIM	0	3	3	l	SMALL	0	SAME
132	VA	500	2	1	a	SMALL	1	DIFF-CLOSE
133	AV	-500	1	1	l	SMALL	0	SAME
134	AV	-500	3	3	l	SMALL	0	SAME
135	AV	-500	8	6	a	LARGE	2	DIFF-CLOSE
136	VA	500	3	9	a	SMALL	6	DIFF-FAR
137	SIM	0	3	2	a	SMALL	1	DIFF-CLOSE
138	VA	500	3	3	l	SMALL	0	SAME
139	VA	500	9	8	a	LARGE	1	DIFF-CLOSE
140	SIM	0	8	1	a	LARGE	7	DIFF-FAR
141	AV	-500	8	1	a	LARGE	7	DIFF-FAR
142	VA	500	1	1	l	SMALL	0	SAME
143	VA	500	9	9	l	LARGE	0	SAME
144	SIM	0	8	8	l	LARGE	0	SAME

Appendix 10: Fraction writing pronunciation

Picture	Number
Flower	½ (read as one-half)
Cat	¼ (read as one-quarter)
Cake	1/7 (read as one-seventh)
Dog	1/3 (read as one-third)
Cow	1/5 (read as one-fifth)
Clown Fish	¾ (read as three-quarters)
Moon	2/5 (read as two-fifths)
Smiley Face	7/10 (read as seven-tenths)
Tick	1 1/8 (read as One and one-eighth)
Cross	2 5/6 (read as Two and five-sixths)