Methods of Timepoint 2 (2018)

Austrian Data

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

Funded by the FWF, Austria, grant number I 2778-G16

Table of Contents

1. Participants

Participants at Time 2 (T2) were 170 children, and they completed all the testing. 169 of these children also had complete data from T1. At T2, 7 children left the study. There was an overall attrition rate from T1 to T2 of 3.95%. Full details of participation and attrition by school and overall can be found in Table 1.

The children were aged between seven and eight years (M_{age}= 97.64 months, *SD* = 3.50 months), 90 males and 80 females, attending Year 2 in Austrian primary schools. Children were recruited from primary schools in the city of Graz.

Testing took place between April and June 2018. The headmasters of all participating schools were informed per e-mail as well as via phone calls. Parents or legal guardians of the children received details about the study in written form and were given the opportunity to opt their child(ren) out of being tested

Children with developmental disorders and/or neurological disorders were not excluded from the study. However, no children with these disorders or special education needs were reported in the Austrian sample.

Table 1: Children at T2: Comparison with T1 and Attrition

School No.	Total Number of Children taking part T1	Number of Children with complete data at T1	Total Number of Children taking part at T2	Total Number of complete data T2	Number of children left at T2	Children with complete data at T1 and T2	Number of children joined at T2	Percentage Attrition (Children left study at T2 divided by total taking part at T1)
1	31	31	31	31	0	31	0	0,00
2	30	30	27	27	3	27	0	10,00
3	24	23	23	23	1	22	0	4,17
4	36	36	35	35	1	35	0	2,78
5	56	56	54	54	2	54	0	3,57
Totals	177	176	170	170	7	169	0	3,95

The study was approved by the ethics committee of the University of Graz (case identification code: 39/23/63 ex 2016/17).

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 2 for tests completed individually and in Table 3 for tests completed during group testing, in the order they were completed.

Table 2. Individual testing: Tests used in the test battery

Cognitive Construct	Test given	Experimental/ Test Origin	Stan-dar-dised	Test Number	Administration time, items, maximum score	Reliability
Number knowledge	Number reading	Experimental	×	Individual test 1	4-minutes (approx.) Total 76 items. Max score 76	Items recoded into 1 and 0, Cronbach's alpha on 76 numbers .97
Number knowledge	Number matching task	Experimental computerised	×	Individual test 3	7 minutes (approx.). 120 items. Reaction time and accuracy data by trial. Max score 120	Bivariate Pearson's Correlation between means for odd and even trials for each participant: r = .92 (p <.001)
Behavioural inhibition	Go-no go task	Experimental computerised	×	Individual test 2	5 minutes. 30 practice items; 80 assessment items (60 go and 20 no-go trials) Reaction time and accuracy data by trial. Max score 80	Bivariate Pearson's Correlation between means for odd and even trials for each participant returned an r =.60 (p <.001)
Reading	Word reading	SLRT-II: Lese-und Rechtschreibtest [Reading and spelling test]. Form A.	✓	Individual test 4	60-seconds per test. Scored as total number of words read correctly in 60 seconds. Items: 156. Max score: 156	Parallel test reliability for Grade 2 = .98 according to manual (Moll & Landerl, 2010).
Reading	Pseudoword reading	SLRT-II: Lese-und Rechtschreibtest [Reading and spelling test]. Form A.		Individual test 5	60-seconds per test. Scored as total number of words read correctly in 60 seconds. Items: 156. Max score: 156	Parallel test reliability for Grade 2 = .96 according to manual (Moll & Landerl, 2010).

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

Cognitive Construct	Test given	Origin	Stan-dar-dised	Test number	Administration	Reliability
Symbolic and non-symbolic magnitude	Magnitude comparison	Experimental	✓	Group test. Session 1, Booklet 1: Practice 1 (symbolic) Practice 2 (non-symbolic) Exercise 1 (symbolic) Exercise 2 (non-symbolic) Star Exercise Group test, Session 1, Booklet 3: 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, 1 star exercise). 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. Star Exercise. 30 seconds. 72 items. Maximum score 72. 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 reliability (including exercises, but not practices or stars) *r* =.90
Number knowledge	Number writing	Experimental	×	Group test. Session 1, Booklet 2. Subtest 1 and 3. 18 items in each subtest. Session 2, Booklet 4, Subtest 1. 19 items. Session 2, Booklet 4, Subtest 7, 19 items. (2 subtests in each session)	Untimed. Approximately 5-minutes per subtest Session 1, Booklet 2, Subtest 1. 18 items. Maximum score 18. Session 1, Booklet 2, Subtest 3. 18 items. Maximum score 18. Session 2, Booklet 4, Subtest 1, 19 items. Maximum score 19. Session 2, Booklet 4, Subtest 7, Maximum scored 19. Scored as total items correct. Also scored at two-digit, three-digit, four-digit, teen correct, and X0, XX, X00, XX0, X0X, XXX. Xteen, X000, XX00, XXX0, X0X0, X0XX, XXXX total correct., number writing total errors, number writing inversions and number writing mirrored digits (total of mirrored digits and by digit mirrored).	Parallel forms reliability (totals for four subtests) *r* =.95 Item by item (on all 74 items) Cronbach's alpha = .98
Mathematical ability	Numerical Operations	Adapted from WIAT II (Wechsler, 2005)	✓	Group Test. Session 1, Booklet 2, Subtest 2	20-minutes, 23 items. Scored as number correct. Maximum score 23.	Including all 23 items alpha = .71 Items 7-23, alpha = .73 (T2)

Cognitive Construct	Test given	Origin	Stan-dar-dised	Test number	Administration	Reliability
Arithmetic fluency	Addition Subtraction	Adapted from Westwood et al. (1974)	✓	Group Test Session 2, Booklet 1: subtests 2: One-minute addition subtest 3: One- minute addition extra, sub-test 4: One-minute subtraction. sub-test 5: subtraction extra.	One minute for each test. One-minute addition. One-minute time limit. 60 items. Maximum score 60. One-minute addition extra. One- minute time limit. 30 items. Maximum score 30. One-minute subtraction. One-minute time limit. 60 items. Maximum score 60. One-minute subtraction extra. One-minute time limit. 30 items. Maximum score 30. Items scored separately by sub-test.	Addition: alpha .91 Addition Extra: alpha .92 Subtraction: alpha .93 Subtraction Extra: alpha .93
Number knowledge	Number identification	Experimental	×	Group Session 2, Booklet 1, subtest 6	16 items, maximum score 16	Cronbach's alpha = .90
Non-symbolic Ordinality	Ordinality task with dots	Experimental	×	Group Session 2, Booklet 1, subtest 8	90 seconds. 80 items. Maximum possible score 80.	
Symbolic (Numerical) Ordinality	Ordinality task (digits)	Experimental	×	Group Session 2, Booklet 1, subtest 9	90 seconds. 80 items. Maximum possible score 80.	

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.

Additional materials for individual testing

The record form of the SLRT-II was used for each child during the individual word and pseudoword reading tasks. Stimuli to be read in the number reading task were shown to children on A4 sheets. Laptops with 15.6inch (Dell Latitude E5570, running Windows 10-resolution 1920 x 1080 pixels) and QWERTZ keyboards were used for the Number Matching and Go-No Go tasks. AKG K 242 HD headphones were used for the Number Matching task. A stopwatch integrated on smartphones was used for administering the reading tasks.

2.1 Transcoding

Number reading. This task was administered during individual testing. This task consisted of 76 numbers in the form of Arabic digits, presented in Calibri (body) font size 20, listed over four A4 pages, with 19 numbers on each page. This task was amended between the first and second time points in the following ways: single digits were removed and five-digit items were added to avoid potential ceiling effects. Twenty-four two-digit numbers, 22 three-digit numbers, 28 four-digit numbers and two five-digit numbers were presented. Two-digit numbers shown on pages one and two were 16, 70, 25, 68, 56, 91, 48, 27, 79, 13, 30, 47, 11, 80, 73, 42, 34, 81, 32 (page 1), 89, 53, 15, 40 and 64 (page 2, as ordered). The three-digit numbers on pages two and three were: 200, 304, 600, 190, 220, 109, 123, 643, 700, 203, 300, 560, 340, 107 (page 2), 242, and 349 (page 3). Four-digit numbers on page 3 were ordered as 8000, 2150, 1015, 2609, 1300, 3791, 1002, and 1060. This was followed by six three-digit numbers: 514, 219, 876, 492, 953, and 538 and three four-digit numbers: 7300, 4500, and 5370. Finally, the following four- and five-digit numbers, displayed on page 4, were 9640, 5346, 8723, 2097, 8043, 3008, 4006, 3802, 9703, 9080, 3050, 9013, 5014, 8012, 7218, 3914, 4615, 23547, and 74532.

Number Writing. This task consisted of 74 numbers in the form of Arabic digits, including 24 double-digit numbers, 22 three-digit numbers and 28 four-digit numbers This test was administered as a part of the group testing and was split across Session one and Session two. The test was administered as 2x18 (Session 1) and 2x19 (Session 2) item blocks (two sub-tests each session). Part 1 on Session 1 consisted of 9 double-digit, 3 three-digit and six four-digit items and part 2 on Session 1 consisted of 3 double-digit, 7 three-digit and 8 four-digit items. The third part

(Session 2) consisted of 9 double-digit, 4 three-digit, and 6 four-digit numbers. The fourth part (Session 2) consisted of 3 two-digit, 8 three-digit, and 8 four-digit items. Number entry was made 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 number writing items in the order they were presented, together with the description of the accompanying pictures can be found in Appendix 1.

Number Identification. This task consisted of two pages with eight rows of numbers on each page, from which the participant selected the correct number (please see Appendix 2 for all numbers used, the target numbers in each row are highlighted). The first two questions contained four items to select from, whilst the remaining questions contained five items each. One single digit, four two-digit, seven three-digit and four four-digit items needed to be identified. 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). At Time 1 testing we used the items used by Göbel et al. (2014), at T2 we added a further eight items (see Table 4).

Table 4: Target Items in the Number Identification test at different time points

Göbel, Watson, Lervåg and Hulme (2014)	**Numer8ED Time 1**	**Numer8ED Time 2**
6	6	6
14	14	14
28	28	28
52	52	52
76	76	76
163	163	163
235	235	235
427	427	427
		514
		123
		643
		349
		2150
		2609
		3291
		7218

Number matching. This computerised task was presented in PsychoPy v1.85.3 (Peirce, 2007). Children were seated at a table in front of a laptop (screen size of 15.6 inches) with a green sticker on the "l" button and red sticker on the "a" button of a standard USB external keyboard. Auditory stimuli were conveyed bilaterally through headphones. Numerical stimuli consisted of digits 1 to 9, except 7, because the number word for 7 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. Number words had different durations depending on their structure: teens had a duration of 0.7s, double-digit numbers had a duration of 1.4s and finally three-digit numbers had a duration of 2s.

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

Sixteen different auditory targets (no teens, no decade numbers, no ties) were presented and each verbal number word was presented four times. To avoid a bias toward "no" responses, on 50% of the trials (*N* = 32) verbal number words were followed by the matching Arabic number. The remaining 32 trials included four possible non-matching distractors, with each distractor-type occurring eight times: (1) an inverted distractor [Unit (U)+Decade(D)+; U+D+], e.g., *twenty-three* → *32,* (2) a distractor where the decade matched with the spoken target number but a different digit appeared in the unit position D+U–, e.g., *twenty-three* → *25,* (3) a distractor where the unit digit of the target appeared at the decade position, whereas the unit position had no overlap U+D–, e.g., *twenty-three* → *35*, and, (4) a nonrelated distractor D–U–, e.g., *twenty-three* → *46*.

The 64 two-digit items were combined with 24 teen numbers and 32 three-digit numbers. Thus, altogether participants were exposed to 120 items (60 matching and 60 non-matching). 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. The same pseudo-randomized sequence was then used for all participants.

2.2 Arithmetic

One Minute Addition and One Minute Subtraction. Each task contained 60 items over two pages in two columns, typed in Calibri (body) font size 24. All questions were single-digit (1 – 9) questions presented in a written format, e.g. 2 + 1 = . The questions began with easier questions (e.g. 2 + 1 = or 2 – 1 =) and gradually increased in difficulty (e.g. 7 + 6 = or 7 – 6 =). Half of the items included carrying on the addition and 20 items of the 60 included borrowing on the subtraction questions**.** For Time 2 a second page of 30 additional items was added to both addition and subtraction subtests. Additional items consisted of more difficult items (including operations on double digits). This test was adapted from Westwood et al. (1974).

One Minute Addition Extra and One Minute Subtraction Extra. In addition, two more subtests were added at Time 2: one *Extra Addition subtest* and one *Extra Subtraction subtest*. These two subtests consisted of thirty items (one page, two columns of calculations). The items were at a higher difficulty level compared to the previous addition and subtraction tasks, and consisted of calculations on two-digit numbers.

Numerical Operations. This test contained 23 questions from the Wechsler Individual Achievement Test 2nd Edition, Numerical Operation subtest (Wechsler, 2005), adapted for group use. Adaptations were: the number 10 was omitted from the number sequence in question three; and questions seven to 15 were changed from being vertical to horizontal in presentation, 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 number of visual objects, with a single-digit total (8).

Questions seven to 23 were formal arithmetic. The children were provided with 15 minutes to complete these formal arithmetic questions after the instructions had been explained. Items consisted of 7 addition items (including three double-digit additions and one three-digit addition), 5 subtraction calculations (including three

double-digit subtraction), 3 multiplication items and two division items. These were set out horizontally as is common practice in this age group, rather than using formal columnar methods.

2.3 Magnitude comparison

These tasks were administered as a part of the group testing during Session 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 (including one practice subtest) and two non-symbolic tasks (including one practice subtest) together with a star and circle comparison task. The first exercise test was a practice comparing digits (symbolic), preceded by two worked through examples. The second exercise was a practice test comparing small squares (non-symbolic comparison) and was also preceded by two worked examples. The two practice trials contained 48 items. This was followed by exercise 'Number 1' (symbolic) and exercise 'Number 2' (non-symbolic), both preceded by two worked examples.

The last task in the first booklet was the 'Star Exercise' where the squares were replaced with stars and circles (72 items). Each pair of items contains one star and one circle, the star was randomly placed either on the left or the right side. Children were required to tick as many of the stars (and none of the circles) as possible in 30 seconds. The Star exercise was added at T2 to provide a measure of general processing speed on the comparison tasks. The stars were 1.3cm high and 1.3 cm wide and the circles were 1.1cm and 1.1cm wide. Both were centrally located in same design as the previous items.

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 3. 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).

Symbolic 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. In the E1 and E4 60 items were presented, one task presented numbers in the close distance (E4) and one task presented numbers in the far distance (E1), 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). These two subtests were matched on problem size item by item.

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 five to 13 dots. For three of the non-symbolic comparison tasks there were 48 items each (E2, E6, E7), display size varied between 20 to 40 dots and the 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). E3 had 60 items and a close distance and E5 had 96 items (increased from 60 at T1) and a far distance between the pairs. Both had an average problem size of 18 and all dots in E3 and E5 were of the same size, and display size varied between five to 13 dots.

Changes from the Time 1 booklet were: the addition of the star exercise and additional items for Exercise 5. Thirty-six additional items (six pages) were added to Exercise 5 to avoid potential ceiling effects if children were able to complete all of the T1 items. To create these additional items the first three pairs were swapped with the last three pairs on a page (for pages 1-6).

2.4 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).

2.5 Go – No Go

This task was computer-administered and designed to measure inhibition. In part 1 children were presented with 30 trials of an image of a cartoon bug and instructed to ‘splat’ it as quickly as they could by pressing the space bar. The inter-stimulus interval varied between 300ms, 600ms and 900ms. The bug stimulus was presented for 500ms. If children responded in less than 500ms, “Splat!” appeared on the screen for 500ms and if there was no response a message reading “Too slow!” appeared for 500ms. Part 2 consisted of 80 trials. Like in part 1, children were instructed to press the space bar to splat bugs, but they were also asked to inhibit their response (not press any keys) when a ladybird was presented. The ladybird was also presented for 500ms. Sixty presentations of the bug

(Go trials) and 20 presentations of the ladybird (No Go trials) were presented in a random order. The inter-stimulus interval varied between 300ms, 600ms and 900ms. The experiment was run using E-Prime Software (version 2.0).

2.6 Word and pseudoword reading

Both the word and pseudoword subtests from the form A of the standardized reading fluency test SLRT-II were used (Moll & Landerl, 2010). For both tasks, the practice trials included two columns of four items each. Test items were disposed on a separate page over eight columns of increasing length and difficulty (four columns with 18 items each and four columns with 20 items each; 156 items overall in each subtest).

3. Procedure

All children took part in two one-hour group testing sessions 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 order of tasks was the same for all children. Individual testing sessions comprised of five tests and group testing sessions comprised of seven tests divided across four Booklets (see Tables 2 and 3).

Group testing sessions

Prior to group testing, children were told that they had to try their best 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 Session 1 Testing

Booklet order:

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

Session 1 Booklet 2: Number writing (Part 1), Numerical Operations, Number Writing (Part 2).

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

Session 1 Booklets 1 and 3

Symbolic and non-symbolic magnitude comparison

Participants were given A5 booklets (two separate booklets were given out during Session one of group testing) containing a total of nine magnitude comparison tasks (including 2 practice tasks) and one star comparison task. 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 total score consisted of the number of correct items for symbolic tasks and the number of correct items for the non-symbolic tasks. In the star task children were given one point for each item in which they correctly ticked the star in the pair (and not the circle). The total score was the number of items with the stars correctly identified. 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, 72 points for the Star Exercise, 60 points for Exercises 3 and 4, 96 points for Exercise 5, and 48 points each for Exercise 6 and 7.

Session 1, Booklet 2

Children were provided with an A4 booklet containing 2 number writing (transcoding) subtests and the 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. Where numbers were written incorrectly, further analysis identified the types of errors made. The maximum possible score of correct items was 72.

All items were coded independently by two researchers in the York team. The initial inter-rater agreement was 93.3% and where there were disagreements as to the appropriate coding this was discussed until agreement was obtained.

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 second part of the task contained seventeen items and children were given up to 15 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. If the child provided the correct answer but with mirrored digits the answer was marked as correct. The maximum possible score was 23. The first six items are number writing and transcoding items, while items from item 7 onwards are arithmetic items. Thus, in addition, for each child we also calculated the number of correct items on this test excluding the first six items (t2nocore, maximum score = 17).

3.2 Session 2 Testing

Booklet order:

Session 2: Booklet 4: number writing part three, one minute addition, one minute addition extra, one minute subtraction, one minute subtraction extra, number identification, number writing part four, ordinality dots, ordinality digits.

Session 2, Booklet 4

One Minute Addition and Subtraction

These tasks comprised of 60 addition items in the one-minute addition sub-test, 30 addition items in the addition extra sub-test, 60 subtraction items in the one-minute subtraction sub-test and 30 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. 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. 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 was 60 for addition, 60 for subtraction, 30 for addition extra and 30 for subtraction extra. Each of the four sub-tests was scored individually.

Number Identification

Children were presented with 16 number identification items. The researcher said a number and the children were required to select it from a row of other 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 for example; 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 the child circled two or more items, they were given a score of 0. The maximum possible score was 16.

Ordinality (dots and digits)

In each task, children were presented with eighty rows, with 10 rows presented on each page, each row contained a triplet of dots or single digits. Items were set out in two columns: Column A and Column B with five items under each column on one page.

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 four 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 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.

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.

Number Reading

Children were shown a list of 76 written numbers in Arabic digit format starting from double-digits going up to five-digits (up to thousands) increasing in length. Numbers were presented over two A4 sheets. Children read out all items.

Coding and scoring. The researcher transcribed children's utterances on a separate marking sheet. One point was given for each number read correctly.

Number matching

This was a computer-based task. 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. 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.

Coding and scoring. The child's responses (reaction time and accuracy) were recorded automatically for each trial.

Go – No Go

This was a computer-based task. The child was shown a bug or a ladybird on the screen and was asked to zap the bug by pressing the SPACE BAR as fast as they could but withhold this response when they saw the lady bird. There were 30 items in the practice session (only zapping the bug) and 80 items in the inhibition task.

Coding and Scoring. Children's accuracy and RT were recorded [accuracy for Go-trials in Part 2 (gngac2g) and No-Go-trials in Part 2 (gngac2ng), RTs for Go-trials in Part 1 (gngmrt1g) and Part 2 (gngmrt2g)].

Word reading fluency

SLRT-II (Moll & Landerl, 2010)

Children were first shown eight practice items of real words over two columns, which the researcher asked the child to read aloud as quickly as possible without errors down the column (starting on the left one). 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 as possible without errors until the experimenters said "stop". The experimenter then turned over the page and asked the child to start reading the words from the top left corner. The number of words correctly read in 60s was recorded.

Coding and scoring. The total number of words read correctly in 60s was recorded. The maximum possible score was 156.

Pseudoword reading fluency

SLRT-II (Moll & Landerl, 2010)

Children were asked to read "fantasy words" (pseudowords). The procedure for practice and test items was the same as the word reading fluency subtest.

Coding and scoring. The total number of words read correctly in 60s was recorded. The maximum possible score was 156.

4. Reference list

Göbel, S. M., Watson, S. E., Lervåg, A., & Hulme, C. (2014). Children's Arithmetic Development: It Is Number Knowledge, Not the Approximate Number Sense, That Counts. *Psychological Science*, *25*(3), 789–798. https://doi.org/10.1177/0956797613516471

Moll, K., & Landerl, K. (2010). *SLRT-II: Lese- und Rechtschreibtest [Reading and Spelling Test]*. Bern, Switzerland: Hans Huber.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8–13.

Wechsler, D. (2005). *Wechsler Individual Achievement Test—Second UK Edition (WIAT-II UK)*. Pearson. https://www.pearsonclinical.co.uk/Psychology/ChildCognitionNeuropsychologyandLanguage/ChildAchievementMeasures/WechslerIndividualAchievementTest-SecondUKEdition(WIAT-IIUK)/WechslerIndividualAchievementTest-SecondUKEdition(WIAT-IIUK).aspx

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

Appendix 1: Number Writing

Session 1

Part1

Picture	Number
Candle	11
Toothbrush	80
Shoes	73
Paint Splat	42
Car	34
Ice-cream	81
Triangle	32
Plane	89
Pencil	53
Bed	700
Ball	203
Hat	1300 (eintausenddreihundert)
Square	953
Fish	3914
Moon	4006
Smiley	9703
Tick	5370
Cross	8012

Part 2

Picture	Number
Leaf	15
Phone	40
Snail	64
Books	300
Chocolate	560
Spoon	340
Apple	107 (einhundertsieben)
Bowl	242
Rubber	349
Lipstick	3791
Hammer	1002 (eintausendzwei)
Ruler	1060 (eintausendsechzig)
Flag	538
Cloud	9013
Pig	7300
Mouse	8723
House	4615
Bike	3802

Session 2

Part 1

Picture	Number
Flower	16
Cat	70
Cake	25
Dog	68
Cow	56
Fish	91
Moon	48
Smiley	27
Tick	79
Cross	200
Star	304
Tree	8000
Bath	514
Car	876
Ice-cream	7218
Triangle	9640
Plane	2097
Pencil	5346
Bed	3050

Part 2

Picture	Number
Chair	13
Table	30
Lollypop	47
Sun	600
Cloud	190
Pig	220
Mouse	109 (einhundertneun)
House	123 (einhundertdreiundzwanzig)
Bike	643
Window	2150
Carrot	1015
Door	2609
Gloves	219
Chocolate	492
Spoon	8043
Apple	3008
Bowl	5014
Rubber	9080
Lipstick	4500

Appendix 2: Number Identification

Which is the right number?

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

i	50014	514	500104	540	541
m	123	10023	100203	132	10032
n	600403	60043	634	643	60034
o	300409	394	300940	30049	349
p	2115	200010050	2150	200150	2000150
r	2006009	2690	2000609	2609	20006009
s	3791	300070091	3000791	3719	3000700901
u	7280	7218	7000218	700020081	700020018

Appendix 3: 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 4: Dot Ordinality

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

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