



Methods of Timepoint 3 (2019)

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	3
2. Materials and Stimuli	4
2.1 Group testing	6
2.1.1 Magnitude comparison	6
2.1.2 Arithmetic	7
2.1.3 Ordinality	9
2.1.4 Numerical knowledge	10
2.2 Individual testing	10
2.2.1 Number matching	10
2.2.2 Conceptual understanding	12
2.2.3 Word and pseudoword reading	12
3. Procedure	12
3.1 Session 1 Testing	13
Symbolic and non-symbolic magnitude comparison	13
Numerical operations	14
One-minute addition and subtraction	14
3.2 Session 2 Testing	15
One-minute multiplication and division	15
Mathematical reasoning	15
Ordinality (dots and digits)	16
Conversion	17
3.3 Individual testing session	17
Number matching	17
Word reading fluency	18
Pseudoword reading fluency	18
Conceptual understanding	19
4. Reference list	19
Appendix 1: Magnitude comparison	20
Appendix 2: Numerical operation	21
Appendix 3: Mathematical reasoning	22
Appendix 4: Dot ordinality	24
Appendix 5: Digit ordinality	26
Appendix 6: Conversion	28
Appendix 7: Conceptual understanding	29

1. Participants

As shown in Table 1, 168 children took part at T3, of which all 168 had complete T3 data. The children all attended Year 3 of Austrian Primary Schools. Of those with complete data, 167 also had complete data from T2, and 166 from T1 and T2. The children at T3 were between eight and nine years old ($M_{age} = 109.06$ months, $SD = 3.47$, $Min = 99$, $Max = 116$). Of the 168 children with complete data, 89 (52,98%) were males and 79 (47,02%) females. The attrition of children from T1 to T3 by school is also shown in Table 1. Testing took place between April and May 2019.

Table 1: Attrition at Time 3

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	N joined at T2	Percentage T2 Attrition (Children left study at T2 divided by total taking part at T1)	Total number of children taking part at T3	Number of children with complete data at T3	Number of children left at T3	children with complete data T1, T2, T3	Number of children joined at T3	Percentage Attrition T1-T3	Percentage Attrition T2-T3
1	31	31	31	31	0	31	0	0.00	30	30	1	30	0	3.23	3.23
2	30	30	27	27	3	27	0	10.00	26	26	1	26	0	3.33	3.70
3	24	23	23	23	1	22	0	4.17	23	23	0	22	0	0.00	0.00
4	36	36	35	35	1	35	0	2.78	35	35	0	35	0	0.00	0.00
5	56	56	54	54	2	54	0	3.57	54	54	1	53	1	1.79	1.85
Totals	177	176	170	170	7	169	0	3.95	168	168	3	166	1	1.69	1.76

Ethics. The project had ethical approval from the University of Graz (case identification code: 39/23/63 ex 2016/17). The headmasters in all participating schools were informed per e-mail as well as via phone calls about the study. Informed written consent was obtained from parents or legal guardians of the children.

2. Materials and Stimuli

We used standardised and non-standardised tasks. Tasks measured a number of cognitive constructs and are detailed in Table 2 for tests completed during group testing and in Table 3 for tests completed individually, in the order they were completed.

Table 2: Group-administered testing: tests used in the test battery at Time 3

Construct	Test	Standardised	Test Session & Number	Administration Time, Items, Maximum Score	Reliability at T3
Symbolic and Non-symbolic magnitude comparison	Magnitude Comparison	Experimental	Session 1, Tests 1 and 3 Session 1 Test 1: Practice 1 (symbolic) Practice 2 (non-symbolic) Exercise 1 (symbolic) Exercise 2 (non-symbolic) Session 1 Test 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) 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 = .87 Parallel forms for non-symbolic = .87 Combined = .91
Mathematical Ability	Numerical Operations	Wechsler Individual Achievement Test (WIAT II) (Wechsler, 2005)	Session 1, Test 2	20 minutes. 26 items. Scored as number of items correct. Maximum score = 26 (note: T1 and T2 items 1-6 were removed from the list for T3).	Cronbach's alpha = .74
	One Minute Addition	Adapted from Westwood, P., Harris-Hughes, M., Lucas, G., Nolan, J., & Scrymgeour, K. (1974).	Session 1, Test 4	One Minute. 60 items. Maximum score = 60.	Cronbach's alpha = .93
	One Minute Addition Extra		Session 1, Test 5	One Minute. 60 items. Maximum score = 60.	Cronbach's alpha = .91
	One Minute Subtraction		Session 1, Test 6	One Minute. 60 items. Maximum score = 60.	Cronbach's alpha = .93

	One Minute Subtraction Extra	One-minute addition test - one-minute subtraction test. Remedial Education, 9(2), 70-72.	Session 1, Test 7	One Minute. 60 items. Maximum score = 60.	Cronbach's alpha = .92
Mathematical Ability	One Minute Multiplication	Adapted from Westwood, P., Harris-Hughes, M., Lucas, G., Nolan, J., & Scrymgeour, K. (1974).	Session 2, Test 1	56 Items. One Minute. Maximum score = 56.	Cronbach's alpha = .91
	One Minute Division		Session 2, Test 2	56 items. One Minute. Maximum score = 56.	Cronbach's alpha = .92
Mathematical Ability	Maths Reasoning	Wechsler Individual Achievement Test (WIAT II)	Session 2, Test 3	Untimed, led by researcher, task took approximately 15 minutes. 19 items. Maximum score = 19.	Cronbach's alpha = .65
Ordinality	Ordinality dots	Experimental	Session 2, Test 4	90 seconds. 80 items. Maximum score = 80.	
	Ordinality digits	Experimental	Session 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)	Session 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 = .82

Table 3. 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 $r = .90, p < .001$
Reading	Word reading	SLRT-II: Lese-und Rechtschreibtest [Reading and spelling test]. Form A.	Individual Test 2	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 3 = .94 according to manual (Moll & Landerl, 2010).
Reading	Pseudoword reading	SLRT-II: Lese-und Rechtschreibtest [Reading and spelling test]. Form A.	Individual Test 3	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 3 = .95 according to manual (Moll & Landerl, 2010).
Mathematical Conceptual Understanding	Mathematical Conceptual Understanding	Experimental	Individual Test 4	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 $r = .41, p < .001$.

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 Session 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 ($3 \times 5 =$). 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. $5 \times 2 =$. The questions began with easier questions (e.g. $9 \times 2 =$) and gradually increased in difficulty (e.g. $7 \times 8 =$). 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 ($25 \div 5 =$). 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. $10 \div 2 =$. The questions began with easier questions (e.g. $8 \div 2 =$) and gradually increased in difficulty (e.g. $72 \div 8 =$). 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 Session 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.

2.2 Individual testing

Additional materials for individual testing

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 7A. The record form of the SLRT-II was used for each child during the individual word and pseudoword reading tasks.

Computer. Laptops with 15.6inch (Dell Latitude E5570, running Windows 10, resolution 1920 x 1080 pixels) or with 14inch (Lenovo Thinkpad, running Windows 10, resolution 1366 x 768 pixels) with QWERTZ keyboards were used for the number matching and the conceptual understanding tasks.

Headphones. AKG K 242 HD headphones were used for the Number Matching task.

Stopwatch. A stopwatch integrated on smartphones was used for administering the reading tasks.

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 7B.

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 7A).

2.2.3 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 displayed 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 three tests and group testing sessions comprised of thirteen tests across 4 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. 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 Session 1 Testing

Booklet order:

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

Session 1 Booklet 2: Numerical Operations, one-minute addition, one-minute addition extra, one-minute subtraction, one-minute subtraction extra

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). 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.

Session 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.

3.2 Session 2 Testing

Booklet order:

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

Session 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.

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 left of the child. The computer and additional keyboard were placed directly in front of the child. 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, Reading (words and pseudowords, SLRT-II), conceptual understanding

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.

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. Children were 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.

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.

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 7A.

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.

4. Reference list

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

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.
<https://doi.org/10.1016/j.jneumeth.2006.11.017>

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
	<i>average</i>	4.85	8.86	5.00	27.05	9.00	4.98	9.02	27.77	29.43
	<i>sd</i>	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
	<i>average</i>	3.29	3.27	5.97	3.48	1.43	1.43	5.97	4.96	8.27
	<i>sd</i>	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	<i>min</i>	0.11	0.38	0.13	0.86	0.78	0.60	0.42	0.82	0.73
	<i>max</i>	0.89	0.92	0.44	0.92	0.91	1.67	0.62	0.88	0.79
	<i>average</i>	0.50	0.70	0.24	0.88	0.85	0.78	0.50	0.84	0.75
	<i>sd</i>	0.24	0.16	0.11	0.02	0.05	0.17	0.06	0.02	0.02
Problem Size	<i>min</i>	3	11	7	43	15	7	15	44	47
	<i>max</i>	17	25	13	64	21	13	21	66	70
	<i>average</i>	9.71	17.73	10.00	54.10	18.00	9.97	18.00	55.54	58.85
	<i>sd</i>	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: 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 14 th ? [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 $\frac{3}{4}$ 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 3 rd of the month, and John's cousin will come to visit on the 17 th , 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

Column 1				Column 2			
Item Number	Number of Dots shown			Item Number	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

Appendix 6: Conversion

Geld

1. € 1 = _____ c
2. € 3 = _____ c
3. 800 c = € _____
4. 670 c = € _____ _____ c
5. € 7 und 5 c = _____ c

Zeit

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

Länge

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

Appendix 7: 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