

# The first steps in developing a number sense screening test for young primary school children

Katja Depolli Steiner\*, Cirila Peklaj, and Anja Podlesek  
Department of Psychology, Faculty of Arts, University of Ljubljana, Slovenia

**Abstract:** Number sense refers to a set of numerical processing skills that develop before entering primary school and evolve with age and experience. Research has shown the importance of these skills for mathematical achievement. Therefore, early identification of students who have difficulty in numerical processing is the key to early intervention to reduce these deficits. This study's purpose was to design a group-administered pencil-and-paper instrument measuring numerical magnitude estimation and numerical magnitude comparison that can be used in the first three grades of primary school as a quick screening tool for number sense deficits. Three quick and easy-to-use tasks measuring non-symbolic and symbolic number sense (the number line estimation task, the area comparison task, the number comparison task) were developed and administered to a group of 316 students in the first three grades of Slovenian primary schools. The results show that these tests provide a good basis for further development of a screening test.

**Keywords:** number sense, numerical processing, symbolic and non-symbolic numerical magnitude, measurement, primary school students

## Prvi koraki v razvoju presejalnega testa za merjenje občutka za števila pri mlajših osnovnošolcih

Katja Depolli Steiner\*, Cirila Peklaj in Anja Podlesek  
Oddelek za psihologijo, Filozofska fakulteta, Univerza v Ljubljani, Slovenija

**Izveček:** Občutek za števila se nanaša na niz veščin numeričnega procesiranja, ki se razvije še pred vstopom v osnovno šolo in se nadalje razvija s starostjo in izkušnjami. Raziskave so potrdile pomembnost teh veščin za matematične dosežke učencev. Zgodnje odkrivanje učencev, ki imajo težave z numeričnim procesiranjem, omogoča zgodnje intervencije in je zato ključnega pomena za zmanjševanje primanjkljajev učencev na tem področju. Namen naše raziskave je bil razviti pripomoček tipa papir-svinčnik za merjenje veščine ocenjevanja in primerjave količin, zasnovan za skupinsko izvedbo v razredu, ki ga lahko uporabimo v prvih treh razredih osnovne šole kot hiter presejalni test za odkrivanje primanjkljajev v občutku za števila. Razvili smo tri naloge za merjenje nesimbolnega in simbolnega občutka za števila (test ocenjevanja na številskih daljicah, test primerjave površin, test primerjave števil); uporaba teh nalog je hitra in enostavna. Teste smo preizkusili na skupini 316 učencev prvih treh razredov slovenske osnovne šole. Rezultati so pokazali, da razviti testi predstavljajo dobro izhodišče za nadaljnji razvoj presejalnega testa.

**Ključne besede:** občutek za števila, numerično procesiranje, simbolna in nesimbolna količina, merjenje, osnovnošolci

\*Naslov/Address: dr. Katja Depolli Steiner, Department of Psychology, Faculty of Arts, University of Ljubljana, Aškerčeva 2, 1000 Ljubljana, e-mail: katja.depolli-steiner@ff.uni-lj.si



Članek je licenciran pod pogoji Creative Commons Priznanje avtorstva-Deljenje pod enakimi pogoji 4.0 Mednarodna licenca (CC BY-SA licenca).  
The article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA license).

In today's increasingly complex and rapidly changing world, the acquisition of basic numerical skills and knowledge is essential for students' successful transition to further education levels and their success in everyday life, e.g., managing their monthly budget or planning significant purchases. However, the data indicate that many students struggle with mathematics in school. Studies of learning difficulties in mathematics show a prevalence of 4 to 6 or 7% among primary school students (Andersson & Östergren, 2012; Kavkler, 2002). In Slovenia, mathematics is one of the most important subjects in primary school. Students must attend at least 4 hours of mathematics per week from grade 1 to grade 9. Beginning in 3rd grade, their achievement is assessed on a 5-point scale from 1 (unsatisfactory) to 5 (excellent). In 2018, 2% of Slovenian 8<sup>th</sup> grade students received grade 1 and 16% received grade 2 in mathematics (Japelj Pavešić & Cankar, 2018), which means that their mathematical skills were very poor according to the national knowledge standards prescribed in the curriculum.

In Slovenia, a slightly adapted Response to Intervention (RTI) model addresses children with learning difficulties. The original RTI model (Fletcher & Vaughn, 2009) is designed for early identification and support of students with learning difficulties. It comprises of three levels (tiers) that allow increasing the intensity of support for students according to their difficulties, is based on a continuous assessment of these difficulties, and involves teachers, school counselors, and parents. Teachers play an essential role in identifying students' difficulties and designing interventions; they can quickly identify and analyze students' difficulties, design an intervention plan, and evaluate it. Teachers can help most students in the classroom by implementing good pedagogical practice (tier 1). For some students (15–20%), this help is not sufficient, and they need additional professional help from counselors (tier 2). However, for about 3–6% of the students, this additional help is still not sufficient, so they need even more intensive special educational and psychological help in the form of intensive individual interventions (tier 3; Fletcher & Vaughn, 2009). In contrast to the original three-tier model, the five-tier RTI model applied in Slovenia (Magajna et al., 2008) includes additional assessment and counseling by school counselors at tier 2, followed by group support by qualified teachers or school counselors (tier 3) and additional diagnostics in a specialized external institution (tier 4). At the last level (tier 5), the adapted RTI model also includes targeted individual interventions.

The key to providing effective help is the early screening of all students who need it. For early identification of students at risk in the general population, we need instruments that enable a reliable and valid identification of students who have problems related to basic numerical skills or representations and help determine which problems are involved. Therefore, in this study, we focused on developing and validating instruments that could be used to identify problems with numerical skills in students in the first three grades of Slovenian primary school. Early identification of students having issues in learning mathematics is important, since early interventions can help prevent such students from

lagging behind their peers (Duncan et al., 2007). Aunio and Räsänen (2016) suggested four main factors for developing core numerical skills: symbolic and non-symbolic number sense, understanding of mathematical relations, counting skills, and basic arithmetic skills. Since developing counting skills and basic arithmetic skills is addressed heavily in the Slovenian mathematics curriculum for Grades 1–3 and is monitored regularly by teachers, we focused on developing an instrument for determining the students' non-symbolic and symbolic number sense that are given less attention in the educational process but are equally important components of mathematical skills.

Number sense is crucial to mathematical achievement (Andrews & Sayers, 2015; Aunio & Niemivirta, 2010; Aunola et al., 2004; Yang & Li, 2008). It refers to the individual's comprehensive understanding of numbers, operations, their relationships, and the ability to deal with situations in daily life where numbers play a role (Yang & Lin, 2015). These skills develop before entering primary school (Aunio & Räsänen, 2016). Longitudinal studies have shown that a number sense measured in preschool at age 4 predicts mathematics achievement 2 years later (Mazzocco et al., 2011). Number sense measured in preschool predicts mathematics achievement in 3rd grade (Jordan et al., 2009, 2010).

Siegler (2016) considers that the ability to distinguish between numerical magnitudes for different numbers (smaller and larger) is crucial for developing the number sense and that the linearity of numerical magnitude representations is also essential. In his integrated theory of numerical development, the knowledge of numerical magnitude, which is reflected in the generation of increasingly precise magnitude representations for an increasingly broad range of numbers, is considered essential for success in the field of mathematics (Siegler, 2016). Magnitude representations first develop on the non-symbolic level (stage 1). The non-symbolic representations are then linked to the symbolic representations for whole numbers (stage 2), followed by an increase in the range of whole numbers whose size an individual can accurately represent (stage 3). The development ends with the extension of these exact representations to rational numbers (Siegler, 2016). For the first three grades of the Slovenian primary school, which are attended by children aged 5–6 to 8–9 years, the first three developmental stages are of crucial importance.

The understanding of number magnitude is based on two inherent cognitive mechanisms. A quantity approximation system (approximate number system – ANS; Dehaene, 2011) is used to represent the approximate value of larger numbers, while a system for determining the exact number (object tracking system – OTS; Andersson & Östergren, 2012; Feigenson et al., 2004) is used to represent small numbers from 1 to 4. ANS allows us to represent the magnitude of numbers using an analogy with a number line. Values on the number line increase from left to right. First, these values are represented logarithmically; children imagine smaller numbers are further apart than large numbers. However, with increasing age and experience, the development of performance on the number line becomes more linear (Praet &

Desoete, 2014). OTS develops very early, even before the age of 1 year (Piazza, 2010); we use it to imagine a small number of objects as separate units through a 1-to-1 correspondence between the object and its mental representation (Andersson & Östergren, 2012; Feigenson et al., 2004). This process is called subitization (Clements, 1999).

Researchers use various instruments to determine an individual's knowledge or representations of numerical quantity and comparisons between numbers. One of the most commonly used instruments to identify symbolic and non-symbolic representations of numerical quantities is the number line estimation task developed by Siegler and his colleagues (Siegler & Booth, 2004; Siegler & Opfer, 2003). The participants are presented with a number line with 0 at one end and 10, 100, or 1000 at the other end, depending on the individual's age (Siegler, 2016). The participants are then shown a number and are asked to indicate where this number is located on the number line. In a complementary task, the participants are shown a position on a number line and are asked to estimate the corresponding number.

Research has shown that representations of whole number magnitudes progress from a consistently logarithmic distribution through a mixture of logarithmic and linear to a primarily linear distribution (Siegler & Booth, 2004; Siegler & Opfer, 2003). Children's representations of small numbers from 0 to 10 pass through this transition between 3 and 6 years of age. This developmental sequence is later repeated for larger numbers, i.e., for numbers from 0 to 100 between 5 and 8 years, for numbers from 0 to 1,000 between 7 and 10 years, and for numbers from 0 to 10,000 between 9 and 12 years (Siegler, 2016). In students with mathematical difficulties or dyscalculia, linear estimation develops later (Geary et al., 2008; Landerl et al., 2009) and is less accurate than that of students without difficulties (Geary et al., 2008; Landerl, 2013).

Studies of representations of number magnitude using number lines have shown that the accuracy of estimating the number magnitude on a number line is related to mathematics achievements, i.e., it is in relation with the accuracy of calculations at the age of 8 years 10 months and 9 years 10 months (LeFevre et al., 2013), with solving word problems at the age of 8 years (Gunderson et al., 2012) and with a standardized math test for students aged 6 to 8 years (Sasanguie et al., 2013).

For the comparison of non-symbolic magnitudes, the task most often used is that of two large squares in which smaller elements (in the form of different objects or signs) are drawn, and children are asked to estimate in which half there are more/fewer elements (Halberda & Feigenson, 2008; Laski & Siegler, 2007). The task difficulty is altered by varying the objects' ratio or the distance between the elements in the sequence. Tasks can be presented on paper or a computer screen. These different tasks are not standardized, and the accuracy of the estimates may also be affected by differences in the visual characteristics of the stimuli (i.e., brightness, distance, size, the area covered by the stimuli), which are difficult to control and may affect the estimates (De Smedt et al., 2013; Gebuis & Reynvoet, 2012). To avoid such problems, we decided to develop an instrument for non-symbolic

magnitude comparison that includes the comparison of the area size instead of the comparison of elements numerosity, since Lourenco and her colleagues (Lourenco & Bonny, 2017; Lourenco et al., 2012) found that non-symbolic cumulative area representations also predict mathematical achievements.

To determine an individual's comparison of symbolic magnitudes, researchers usually use tasks in which two numbers written in Arabic numerals are shown and participants are asked to estimate which of the two is larger/smaller in magnitude. The numbers used in these tasks can be small, e.g., between 1 and 9 (De Smedt et al., 2009), or can be chosen from a broader range of orders of magnitude, e.g., between 1 and 100, if the participants are older students (Kolkman et al., 2013). The numbers can be displayed on a computer screen (Sasanguie et al., 2012) or read aloud by the experimenter if this task is applied to younger children (Laski & Siegler, 2007). Accuracy, response time, or numerical distance/ratio effect can be recorded (De Smedt et al., 2013).

Studies of normative population students have shown that mathematical achievement is related to response accuracy (De Smedt et al., 2009; Kolkman et al., 2013), reaction time (De Smedt et al., 2009; Sasanguie et al., 2013), the adjusted score of response speed and accuracy (Sasanguie et al., 2012, 2013), the distance effect (De Smedt et al., 2009) and the ratio effect (Sasanguie et al., 2013). Research has also shown that students with dyscalculia aged 6 to 10 years attain poorer magnitude estimation results than a normative population. These poor results are reflected in response accuracy (Rousselle & Noël, 2007) and reaction times (De Smedt & Gilmore, 2011; Landerl & Kölle, 2009; Landerl et al., 2004). Since the research results show a significant correlation between students' ability to compare number magnitudes and their further achievement in mathematics, we have decided to include in our instrument a task of comparing the magnitude of two numbers.

The main purpose of our study was to develop a convenient instrument for screening students with problems in number sense development at the beginning of primary school. We focused on two components of number sense, i.e., numerical magnitude estimation and magnitude comparison, both on the non-symbolic and symbolic levels. Studies (e.g., Booth & Siegler, 2008; Geary, 2011; Schleeper et al., 2016; Xenidou-Dervou et al., 2017) have shown that these two skills are the key numerical competencies related to students' continued mathematics achievement. They have also shown that students with learning difficulties in mathematics and dyscalculia have difficulties in developing these skills (e.g., De Smedt & Gilmore, 2011; Landerl & Kölle, 2009; Landerl et al., 2004; Rousselle & Noël, 2007). Early identification of students who have difficulty in numerical processing is key to early intervention to reduce these deficits (e.g., Gersten et al., 2005). We aimed to develop an instrument that can be used by school psychologists as a screening instrument. To make the use of such instrument as convenient and cost-effective as possible, we decided to design a set of paper-and-pencil tasks that are short and can be applied in groups in the classroom. In this article we present the first steps in the development of this instrument.

## Method

### Participants

A total of 316 students (166 boys and 150 girls) from five Slovenian primary schools participated in the study. Of these students, 117 were in the first grade (62 boys), 115 in the second grade (63 boys), and 84 in the third grade (41 boys). Their average age was about seven years ( $M = 83.4$  months,  $SD = 3.78$ ), eight years ( $M = 95.8$  months,  $SD = 4.20$ ), and nine years ( $M = 107.3$  months,  $SD = 3.75$ ), respectively.

### Procedure

We obtained parental consent for all students that participated in the study. The testing took place in school classrooms in small groups (approx. 10 to 15 students). Each group was tested in one session during regular school hours. Typical duration of the testing as a whole was 45 minutes, including material distribution and collection, and short breaks (as needed by specific groups of students). The testing was conducted by an experimenter (one of the researchers) and her assistant (a 5th-year psychology student). For all participating students, we also obtained teachers' assessment of their mathematical skills.

### Instruments

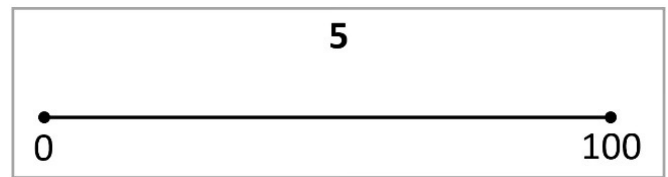
**The number sense test.** The developed number sense test was composed of three tasks: the number line estimation, area comparison, and number comparison task.

In the *Number line estimation task*, two 10-item sets, i.e., number-to-position (NP) and the position-to-number (PN) item set, were adapted from Siegler and Opfer (2003). Each item consisted of a 10 cm long line with left end labeled "0" and right end labeled "100". One randomly selected number from each interval of ten was used (0–10, 10–20, and so forth to 90–100). The number lines were printed in a scrambled arrangement on a landscape page in A3 format (five tasks on the left side of the page, five tasks on the right side of the page). The test pages were preceded by a separate page with two sample items to ensure that the students understood the task and knew the interval size. After the sample items were presented, the students went through the task at their own pace, with the time for the whole set limited to two minutes. A time limit was set to keep testing time within reasonable limits.

In Set A consisting of the NP items, the non-symbolic numerical magnitude estimation was measured by asking students to locate a given number on a 0–100 number line (see Figure 1). The experimenter's instructions were the following: "What I am going to ask you to do is to mark the position of some numbers on the number line. Our number line goes from 0 at the left end to 100 at the right end. Where would you put 5? Mark it on a line like this [making a vertical hatch mark]". Set B consisting of the PN items measured the symbolic numerical magnitude estimation. Students were asked to estimate the number corresponding to a marked position on a 0–100 number line (see Figure 2).

**Figure 1**

*Sample Item for Set A of the Number Line Estimation Task*



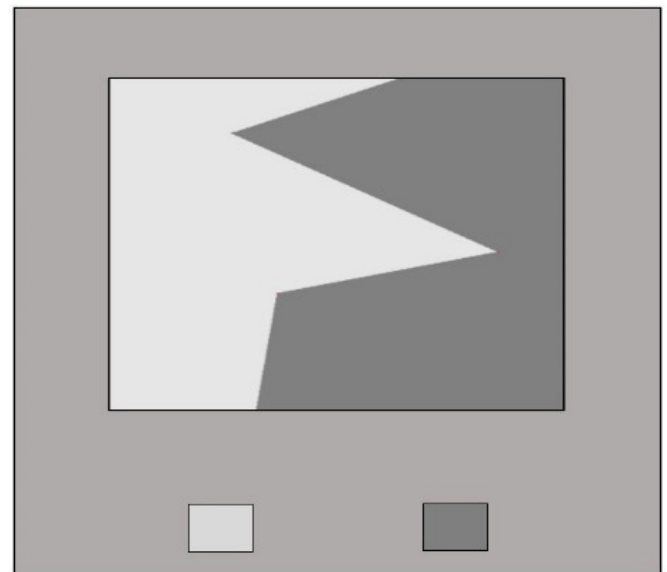
**Figure 2**

*Sample Item for Set B of the Number Line Estimation Task*



**Figure 3**

*One of the Items in the Area Comparison Task, Ratio 11:10 (The Dark Gray Part Is Larger)*



The experimenter's instructions were: "Now I am going to ask you to decide which numbers are already marked on the number line. Our number line goes from 0 at the left end to 100 at the right end. What is this number [pointing to the hatch mark]? Write it here [pointing to the short line above the number line]." The content of different items is listed in Table A1 in appendix.

The *Area comparison task* measured the non-symbolic numerical magnitude comparison. As there was no suitable precedent in the literature, this task was created for the purposes of this study. The items were in the form of a square divided into two areas, one colored light gray and the other dark gray. Students had to estimate which of the two areas is larger (see Figure 3). The instructions of the experimenter

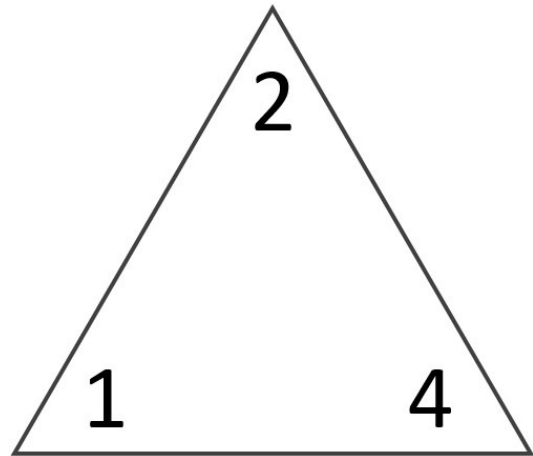


were the following: “Look at the square. It consists of two parts, one dark gray and one light gray. Which part is bigger? When you decide, I want you to put a checkmark in the box of the same color under the square”. Three sets of items (sets A, B, and C) were designed; each set covered ratios 6:5, 7:6, 8:7, 9:8, 10:9, and 11:10, with four items for each ratio, so that all possible combinations of color (light gray, dark gray) and position (left, right) of the larger area were included (see Table A2 in appendix). Within each set, the items were randomly ordered and presented in three rows of three items each on eight A4 pages, starting with set A, followed by set B and set C. The test pages were preceded by a separate page with three sample items to ensure that the students understood the task. After presenting the sample items, the students went through the task at their own pace, and the total time was limited to five minutes. A time limit was set to keep testing time within reasonable limits.

The *Number comparison task*, which is modelled on one of the tasks in the Number Knowledge Test (McGraw-Hill Education, n.d.), measured the symbolic numerical magnitude comparison. It consisted of 24 increasingly difficult items (Figure 4) presented in three rows of two on four A4 pages. The students had to decide quickly without calculations which of the two numbers in the lower corners of an equilateral triangle (i.e., the comparison number) was closer in magnitude to the number in the upper corner (i.e., the reference number). Comparison numbers were chosen so that their distances from the reference number either differed by 1 (in 12 items) or by 2 (in the other 12 items). For example, in the triangle with the reference number 3 where the two comparison numbers were 1 and 7, the distances (i.e., the absolute difference) between the reference and comparison numbers were 2 and 4, so the difference between the distances was 2. In items 1–6, the reference numbers had values between 3 and 8, and if the comparison to the other two numbers was done through performing calculations, the arithmetic would require no carrying over. In items 7–12, the reference numbers were larger (13–18) and no carrying over would be required when making comparisons. In items 13–18, the reference and comparison numbers had values up to 20 and one of the comparison numbers would require a carry over or borrowing. In items 19–24, the reference numbers were larger than 20 and a carry over or borrowing would be needed for one of the comparisons. Items are shown in Table A3 in appendix. The experimenter’s instructions were: “Now I am going to ask you to compare some numbers. Here is a triangle. First, look at the number in the upper corner. Now, look at the numbers in the two lower corners. Which one of the numbers in the lower corners is closer in magnitude to the number in the upper corner? Do not make any calculations; choose the number you think is right and circle it.” To prevent performing calculations, the experimenter guided the pace by reading the numbers in the upper corners at intervals of three seconds. Students were also given instructions on what to do if they got lost due to temporary inattention: “If you miss a triangle, just wait for the next one.” The test pages were preceded by a separate page with three sample items and a practice page with six items to ensure that all students understood both the task and the guided solving. At the end

**Figure 4**

*Sample Item for the Number Comparison Task*



of each test page, the testator stopped and made sure that all students had managed to turn the page and were ready before continuing with the task.

**Other measures.** For the validation of our newly designed instrument, we used teacher’s assessment of students’ numerical skills scale and the test of writing two- and three-digit numbers.

*Students’ numerical skills scale* consists of 11 items for assessing following skills of the students on a 7-point scale (from 1 – very poor to 7 – excellent): (1) speed in solving tasks, (2) correctness of answers, (3) speed in retrieving mathematical facts (e.g., addition/subtraction up to 10, times-tables), (4) use of adequate mathematical procedures (e.g., addition/subtraction, subtraction of two digits with borrowing), (5) use of age-appropriate calculation strategies (e.g., counting with fingers), (6) autonomy in solving mathematical problems, (7) persistence in solving mathematical problems, (8) interest in mathematics, (9) mathematical knowledge, (10) reading skill, and (11) writing skill. Most items relate to core numerical skills, except for the last two items, which relate to reading and writing. Reading and writing are also crucial for solving mathematical tasks.

All 11 items were included in a principal component analysis. A scree plot showed a one-component solution. The first component explained 79.45% of the total variance. For each student, teacher’s answers to 11 items were averaged to obtain the scale score. The scale showed a strong internal consistency ( $\alpha = .97$ ).

In the *writing numbers test*, the experimenter read out loud numbers and students wrote them down on a response sheet. Part 1 comprised 10 two-digit numbers and part 2 comprised 10 three-digit numbers. Students in the second and third grade completed both parts, while students in the first grade only completed part 1 as they were not yet familiar with three-digit numbers. For each student, the correct answers in each part were counted to obtain the test score. Both parts showed a strong internal consistency ( $\alpha = .92$  for part 1 and  $.97$  for part 2).

## Results

For each item in the number line estimation tasks, the absolute error (the deviation from the correct answer) was used as the student's score. The mean scores for each item, item-total correlation, and Cronbach's alpha if the item was deleted are shown in the appendix (Table A1). The student's score for the task was calculated as the mean absolute error for all items included; a higher number means a lower score.

For the area comparison task and the number comparison task, all test items were scored as 1 for the correct answer and 0 for the incorrect answer or the unanswered item. The percentage of students who answered each item correctly, item-total correlation, and Cronbach's alpha if the item was deleted are shown in the appendix (Tables A2 and A3). The student's score for the task was calculated as the sum of the correct answers for all items included in the task; a higher number means a higher score.

*Number line estimation tasks.* Overall, both parts of the number line estimation task were easy for the third grade students, reasonably easy for the second grade students, but quite difficult for the first grade students. In the NP task, the average score of the students in third grade was 7.17, in second grade 10.10 and first grade 17.93. The students' average scores in the PN task were 6.71, 9.36 and 19.35, respectively. For all items, the average absolute error decreased from the first to the third grade, which shows that the students' performance increased with the grade. In the NP task, the average absolute error for the individual items ranged from 12.66 to 3.48 in the third grade, 15.25 to 5.39 in the second grade, and 13.23 to 2.42 in the first grade. In the PN task, the average absolute error for the individual items ranged from 10.90 to 0.83 in the third grade, 24.88 to 11.73 in the second grade, and 29.44 to 6.63 in the first grade. A closer examination of the average absolute errors for individual items showed that the two items with numbers closest to 0 (numbers 3 and 6) were the easiest in both tasks. The PN item with the number 98 had a high accuracy as well. Only one NP item (number 86) and two PN items (numbers 74 and 67) had an average absolute error of more than 10 in all three grades. These results show that representations close to reference points 0 and 100 are easier to make than those further away from such points.

Both parts of the number line estimation task showed acceptable internal consistency in all three grades ( $\alpha$  were .79, .81 and .68 for the NP task and .81, .80 and .73 for the PN tasks). Detailed psychometric information for individual items on both number line estimation tasks are listed in Table A1 in the appendix. In both the NP and PN tasks, all items were retained.

*Area comparison task.* During this task, the experimenters noticed that students' attention and motivation quickly decreased. There were some cases where students did not get to the last item because of a time limit. After testing, it was therefore decided to reduce the length of the test from three to two sets of items for analysis and future use. Items with highest accuracy (48 items) were chosen from the three original sets and 24 items were removed. The items from sets A and B were preferred because they were reached and solved (either correctly or incorrectly) by a larger number

of participants. Therefore, only four items from set C were retained, as a comparable substitute for three items from set A and one item from set B. The selected four items from set C appeared to perform better than the replaced items in sets A and B. The retained items analysis is shown in Table A2 in appendix. The final version of the task showed an acceptable internal consistency in all three grades ( $\alpha = .80, .72$  and  $.68$ ). Overall, the task was relatively easy – in the third grade, 37 retained items were answered correctly by more than 70% of the students (33 items in the second grade, 31 items in the first grade), 11 items were answered by 50 to 70% (14 items in the second grade, 12 items in the first grade), and none by less than 50% (one item in the second grade, five items in the first grade). A closer examination of response accuracy showed that it was highest in items with 6:5 ratio (82.9, 85.8 and 85.1%) and decreased with higher ratios. Based on these results we estimate that for children in the first three grades of primary school, the just noticeable difference in surface area is roughly at the ratio of 9:8 (at this ratio the probability of correct response exceeded the chance level by approx. 50%).

*Number comparison task.* This task was relatively easy, with 19 of the 24 items answered correctly by more than 70% of the third grade students (16 items in the second grade, six items in the first grade), and another two items answered correctly by 50 to 70% of the third grade the students (five items in the second grade, 12 items in the first grade). Only three items, items 19 to 21, were shown to be hard, as they were answered correctly by only approx. one-third of the students. All 18 items with numbers under 20 were answered correctly by more than 50% of the students in the second and third grades. The easiest items were those involving numbers under 10. Most difficult were those involving larger numbers (e. g., Item 21, see Table A3 in appendix). Overall, the items with a difference of 2 between the distances of the comparison numbers to the reference number were easier than those with a difference of 1. All items in this task were retained. The task showed acceptable internal consistency in all three grades ( $\alpha = .63, .62$  and  $.77$ ).

*Differences between grades.* The next step was to compare the performance of students from different grades. Four one-way independent ANOVAs were conducted to compare group mean scores on the number line estimation, area comparison, and number comparison tasks. Table 1 shows descriptive statistics for the overall sample and separately for the first, second, and third grade students. The results from ANOVA are listed in the last column of the table. The mean scores' differences were statistically significant in all tasks, with medium to large effect sizes. The mean scores of the three groups consistently showed that students' performance increased with grade. The largest increase in mean scores from the first to the third grade, with effect sizes ( $\omega^2$ ) of .26 and .35, was found in the two number line estimation tasks. The smallest increase was found in the area comparison test, with an effect size ( $\omega^2$ ) of .05.

*Correlations with teacher's assessment and writing numbers.* To assess criterion validity of the tasks, we calculated Pearson correlation coefficients between test scores and other measures of students' numerical skills, separately for the first, second, and third grade. The results are shown in Table 2.

**Table 1**

*Means (and Standard Deviations in Parentheses) of the Test Scores in the Overall Sample and the Comparison of Students from Different Grades*

Task	Overall ( <i>n</i> = 316)	1 <sup>st</sup> grade ( <i>n</i> = 117)	2 <sup>nd</sup> grade ( <i>n</i> = 115)	3 <sup>rd</sup> grade ( <i>n</i> = 84)	Results of ANOVA		
					<i>F</i> (2, 313)	<i>p</i>	$\omega^2$
Number line estimation							
NP task	12.01 (8.66)	17.93 (10.07)	10.10 (6.66)	7.17 (3.52)	53.94	< .001	.26
PN task	12.09 (9.11)	19.35 (10.54)	9.36 (5.89)	6.71 (3.00)	80.59	< .001	.35
Area comparison	36.41 (5.62)	34.79 (6.21)	36.88 (5.18)	38.04 (4.72)	9.27	< .001	.05
Number comparison	16.00 (4.12)	14.38 (4.27)	16.08 (3.50)	18.15 (3.69)	23.60	< .001	.12

Most statistically significant correlations were found between the number sense test scores and the teacher's assessment, indicating a positive relationship between the students' higher test results and the teacher's higher assessment of their skills. Both number line tasks showed moderate and statistically significant correlations in all three grades, and similar results were found for the number comparison task. In contrast, correlations between the teacher's assessment and the area comparison task were statistically non-significant in the first and second grades. In the third grade this correlation reached statistical significance but was weak. It should be noted, however, that the differences between the correlation coefficients between grades were not statistically significant, with the exception of the correlations between the teacher's assessment and the number comparison task in the first and second grade ( $z = 1.88, p = .03$ ).

The correlations between the number sense test scores and the scores for writing two- and three-digit numbers also indicated a positive relationship between the students' higher test results on some tasks and students' better skill in writing two- and three-digit numbers. Writing two-digit numbers correlated significantly only with the number line tasks and the number comparison task in the first grade; the correlations were moderate. The differences between these correlation coefficients between grades were statistically significant only for the correlations between writing two-digit numbers and the two number line task in the first and second grade (NP task:  $z = -3.49, p < .001$ ; PN task:  $z = -4.18, p < .001$ ) and in the first and third grade (NP task:  $z = -1.88, p = .030$ ; PN task:  $z = -2.86, p = .002$ ). Writing three-digit numbers also correlated significantly only with the number line tasks and the number comparison task in the second grade; the correlations were moderate. The differences in correlation coefficients between grades were statistically significant for both number line tasks (NP task:  $z = -2.46, p = .007$ ; PN task:  $z = -2.59, p = .005$ ) and for the number comparison task ( $z = 2.26, p = .01$ ).

## Discussion

Our study aimed to develop a convenient instrument for screening students with problems in numerical skills at the beginning of primary school. In this article we presented the first steps in the development of this instrument. We designed

three tasks to assess two numerical processing skills that are part of the number sense, i.e., numerical magnitude estimation and numerical magnitude comparison, both on the non-symbolic and symbolic level. The non-symbolic numerical magnitude processing is covered by the number line estimation task (NP task) and the area comparison task, while the symbolic numerical magnitude processing is covered by the number line estimation task (PN task) and the number comparison task. Our results indicate that the test provides a good basis for further development, as it has sufficient initial psychometric characteristics. The internal consistency of the tasks is acceptable. However, in further development of the test, it is important to ensure a much larger sample for all three grades.

The first confirmation of the test's validity was the finding that the students' performance in all three tests increased with grade level, which is consistent with the finding that number sense develops with age (e.g., Halberda & Feigenson, 2008; Siegler, 2016). The test's validity was additionally confirmed by linking its scores to other measures of students' numerical skills, i.e., teacher's assessment and writing numbers. Even though not all correlations were significant for the individual grade levels, the directions of the correlations were as expected.

The relationship between teachers' assessment of students' numerical skills and test scores was confirmed for the NP, PN, and the number comparison tasks, but not for the area comparison task. Teachers mainly assessed students' basic arithmetic skills, as the math curriculum in the first three years of primary school in Slovenia focuses on addition/subtraction, the use of basic mathematical procedures, and related problems. Geometry, where non-symbolic magnitude representations could have a significant influence, is not yet an important part of the curriculum in these grades.

As for our second measure of students' numerical skills (writing numbers), only associations between writing two-digit numbers and test scores in the first grade and associations between writing three-digit numbers and test scores in the second grade were found. These results may also be explained with the math curriculum in the first three years. For first-grade students, writing two-digit numbers is a new and not yet consolidated task, but it should not be difficult for second- and third-grade students. Similarly, writing three-digit numbers is a new task for second-grade students, but not for third-grade students. It is also not surprising that writing

**Table 2**  
*Correlations Between the Number Sense Test Scores and Criterion Scores, and Pairwise Correlations Comparisons*

Task	1 <sup>st</sup> grade (n = 117)	2 <sup>nd</sup> grade (n = 115)	3 <sup>rd</sup> grade (n = 84)	Pairwise comparison	
				1 <sup>st</sup> grade - 2 <sup>nd</sup> grade	1 <sup>st</sup> grade - 3 <sup>rd</sup> grade
Correlations with teacher's assessment of students' numerical skills					
Number line, NP task	-.30**	-.41**	-.49**	z = 0.94, p = .174	z = 1.56, p = .059
Number line, PN task	-.30**	-.46**	-.34**	z = 1.41, p = .079	z = 0.31, p = .378
Area comparison	.16	.17	.23*	z = -0.08, p = .468	z = -0.50, p = .309
Number comparison	.50***	.29***	.44***	z = 1.88, p = .030	z = 0.53, p = .298
Correlations with score in writing two-digit numbers					
Number line, NP task	-.40**	.04	-.15	z = -3.49, p < .001	z = -1.88, p = .030
Number line, PN task	-.49**	.02	-.12	z = -4.18, p < .001	z = -2.86, p = .002
Area comparison	.12	.12	.05	z = 0.00, p = .500	z = 0.49, p = .312
Number comparison	.34***	.13	.17	z = 1.76, p = .039	z = 1.26, p = .104
Correlations with score in writing three-digit numbers					
Number line, NP task	/	-.37**	-.03	/	z = -2.46, p = .007
Number line, PN task	/	-.42**	-.07	/	z = -2.59, p = .005
Area comparison	/	.00	.11	/	z = -0.76, p = .224
Number comparison	/	.30**	-.02	/	z = 2.26, p = .012

*Note.* / = 1<sup>st</sup> grade students did not complete this task.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .



two and three-digit numbers is not correlated with the area comparison task score in any grade, as this task measures the non-symbolic magnitude representations.

To enable universal, quick application of the test in the classroom, we have designed it as a group-administered pencil-and-paper measure. Our study showed that all tasks used in the tests were relatively easy to understand and complete. Even the youngest students in the first grade had no problems understanding the experimenter's instructions and performing the tasks. Time limits were selected properly. Nevertheless, the groups should not exceed ten students to ensure a proper setting and allow the experimenter to assist all students when applying the test.

The best measure among the tasks was the number line estimation task. It has acceptable internal consistency, shows moderate correlations with the other two measures of students' numerical skills, and shows the largest increase in students' performance from the first to the third grade, with medium effect size. These findings are consistent with previous research showing that the number line estimation task is a robust tool for predicting mathematical competence. For example, in their meta-analysis, Schneider et al. (2018) reported an average correlation of 0.44 between number line estimation abilities and broader mathematical competence. However, the test tasks were identical for all three groups and did not account for any possible developmental differences between the students that could affect their performance. For example, a larger format of test materials is generally used for younger students, e.g. a 25 cm long line on separate pages (see Siegler & Opfer, 2003).

The number comparison task also proved to be a good measure, with acceptable internal consistency, moderate correlations with the two measures of students' numerical skills, and a significant increase in students' performance with grade. Symbolic numerical magnitude comparison has previously been associated with young students' numerical abilities, such as fourth-graders' multiplication fact ability (Schleepen et al., 2016) or second-graders' general mathematics achievement (Xenidou-Dervou et al., 2017).

The weakest measure seems to be the area comparison task. It shows an acceptable internal consistency and a slight increase in students' performance with grade. Nevertheless, it shows only a weak correlation with teacher's assessment in the third grade students. We feel, however, that it is reasonable to retain this task in the test and further develop it for two reasons. First, there is some, albeit limited, research evidence of a correlation between non-symbolic magnitude comparison and mathematical ability (for example, Lourenco & Bonny, 2017; Schleepen et al., 2016, Xenidou-Dervou et al., 2017). Second, this task could help identify students with a severe deficit in non-symbolic numerical magnitude representations. These students are at risk of having problems with geometry when it becomes part of the curriculum. As Clements and Sarama (2011) have pointed out, the domain of geometry and spatial reasoning is an important area of mathematics learning, and education in geometry may contribute to a growth in mathematical competence. However, as this task was newly created, it should be further developed in different variants of forms in order to find the most appropriate one for the students of the target age.

*Limitations and directions for future research.* When interpreting our results, some limitations should be considered. First, we report on the first steps of the development of the new instrument. The developmental differences between the students have not yet been fully considered. This should be subsequently changed by further developing the tasks. For example, the number line estimation tasks could be presented on separate pages and a larger format could be used for the younger students.

Second, further investigations of its psychometric properties with improved tasks, other samples, and the application of other measures (e.g., of executive functions) for its further validation are required. A thorough standardization of the instrument would allow for its wider use as a screening and diagnostic tool.

Third, although the same number of classes from the three grades was included in our sample, the final sample of third-grade students was considerably smaller than the first and second-grade students' samples. For reasons unknown to us, the proportion of parents who did not consent to their children's participation in the study was higher in the third than in the first and second grades. Unequal sample sizes could influence the results, so it is essential to ensure that future studies are carried out with more similar sample sizes.

Fourth, our results suggest that instruments measuring number sense should focus on symbolic numerical processing skills, as these appear to be more strongly associated with numerical skills than non-symbolic processing skills. This finding is consistent with other research; for example, Schneider et al. (2018) also found in their meta-analysis that the association with mathematical skills is stronger for symbolic than for non-symbolic numerical magnitude processing. However, since geometry, in which non-symbolic magnitude representations could have a significant influence, is also an essential part of mathematics, we believe that measures of these representations, including the area comparison test, which has not been shown as a strong predictor of mathematical achievement, should remain included in our test. Further work is needed to validate the test on older students and with other measures that include various non-symbolic tasks. It would also be interesting to look for gender differences.

*Conclusions.* The developed number sense test proved to have potential to become a quick, sufficiently reliable, and easy-to-use screening test to measure both the non-symbolic and symbolic number sense in students at the beginning of primary school. At this stage, however, the test is not yet suitable for use, but it provides a good basis for further development. In the next steps, the test will need to be developed and validated further, and norms will have to be developed in order for the test to be used as a diagnostic instrument. Computerized application should probably also be considered. Our test was developed and tested before the Covid 19 epidemic, when the use of computers in working with young students was not widespread in Slovenian primary schools. With the epidemic, this has changed, many aspects of schoolwork have been digitized, schools have gained more ICT, students are more experienced in using it, and therefore computerized use of the test should be fairly straightforward.

Early intervention is of great importance for students who have difficulties in numerical processing. It enables their teachers and school counselors to address these deficits to reduce the students' risk of low mathematics achievement. In Slovenia's five-tier RTI model, our test, when further developed, could be used at tier 2 as a screening tool for the first three primary school grades, helping select the students eligible for group support at tier 3. Targeted interventions could then be designed to help the selected students overcome their deficits in number sense, perhaps in the form of computer-assisted practice as it was already found in previous studies (Park & Brannon, 2013; Wilson et al., 2006).

## Acknowledgment

The authors acknowledge the financial support from the Slovenian Research Agency (research core funding No. P5-0110).

## References

- Andersson, U., & Östergren, R. (2012). Number magnitude processing and basic cognitive functions in children with mathematical learning disabilities. *Learning and Individual Differences*, 22, 701–714. <https://doi.org/10.1016/j.lindif.2012.05.004>
- Andrews, P., & Sayers, J. (2015). Foundational number sense: A framework for analysing early number-related teaching. In O. Helenius, A. Engström, T. Meaney, P. Nilsson, E. Norén, J. Sayers, & M. Österhom (Eds.), *Developments of mathematics teaching: Design, scale, effects* (pp. 17–26). SMDF.
- Aunio, P., & Niemivirta, M. (2010). Predicting children's mathematical performance in grade one by early numeracy. *Learning and Individual Differences*, 20(5), 427–435. <https://doi.org/10.1016/j.lindif.2010.06.003>
- Aunio, P., & Räsänen, P. (2016). Core numerical skills for learning mathematics in children aged five to eight years – A working model for educators. *European Early Childhood Education Research Journal*, 24(5), 684–704. <https://doi.org/10.1080/1350293X.2014.996424>
- Aunola, K., Leskinen, E., Lerkkanen, M. K., & Nurmi, J.-E. (2004). Developmental dynamics of math performance from preschool to Grade 2. *Journal of Educational Psychology*, 96(4), 699–713. <https://doi.org/10.1037/0022-0663.96.4.699>
- Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development*, 79(4), 1016–1031. <https://doi.org/10.1111/j.1467-8624.2008.01173.x>
- Clements, D. H. (1999). Subitizing: What is it? Why teach it? *Teaching Children Mathematics*, 5, 400–405.
- Clements, D. H., & Sarama, J. (2011). Early childhood teacher education: The case of geometry. *Journal of Mathematics Teacher Education*, 14(2), 133–148. <https://doi.org/10.1007/s10857-011-9173-0>
- Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. Oxford University Press.
- De Smedt, B., & Gilmore, C. K. (2011). Defective number module or impaired access? Numerical magnitude processing in first graders with mathematical difficulties. *Journal of Experimental Child Psychology*, 108(2), 278–292. <https://doi.org/10.1016/j.jecp.2010.09.003>
- De Smedt, B., Noël, M.-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, 2(2), 48–55. <https://doi.org/10.1016/j.tine.2013.06.001>
- De Smedt, B., Verschaffel, L., & Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *Journal of Experimental Child Psychology*, 103(4), 469–479. <https://doi.org/10.1016/j.jecp.2009.01.010>
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428–1446. <https://doi.org/10.1037/0012-1649.43.6.1428>
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8(7), 307–314. <https://doi.org/10.1016/j.tics.2004.05.002>
- Fletcher, J. M., & Vaughn, S. (2009). Response to intervention: Preventing and remediating academic difficulties. *Child Development Perspectives*, 3(1), 30–37. <https://doi.org/10.1111/j.1750-8606.2008.00072.x>
- Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, 47(6), 1539–1552. <https://doi.org/10.1037/a0025510>
- Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology*, 33(3), 277–299. <https://doi.org/10.1080/87565640801982361>
- Gebuis, T., & Reynvoet, B. (2012). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General*, 141(4), 642–648. <https://doi.org/10.1037/a0026218>
- Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities*, 38(4), 293–304. <https://doi.org/10.1177/00222194050380040301>
- Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2012). The relation between spatial skill and early number knowledge: The role of the linear number line. *Developmental Psychology*, 48(5), 1229–1241. <https://doi.org/10.1037/a0027433>
- Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the “number sense”: The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, 44(5), 1457–1465. <https://doi.org/10.1037/a0012682>

- Japelj Pavešić, B., & Cankar, G. (2018). Linking mathematics TIMSS achievement to national examination scores and school marks: Unexpected gender differences in Slovenia. *Orbis Scholae*, 12(2), 77–100. <https://doi.org/10.14712/23363177.2018.294>
- Jordan, N. C., Glutting, J., Ramineni, C., & Watkins, M. W. (2010). Validating a number sense screening tool for use in kindergarten and first grade: Prediction of mathematics proficiency in third grade. *School Psychology Review*, 39(2), 181–195. <https://doi.org/10.1080/02796015.2010.12087772>
- Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, 45(3), 850–867. <https://doi.org/10.1037/a0014939>
- Kavkler, M. (2002). Kako otroci rešujejo osnovne matematične probleme [How children solve mathematical problems]. In N. Končnik Goršič & M. Kavkler (Eds.), *Specifične učne težave otrok in mladostnikov* (pp. 157–172). Svetovalni center za otroke, mladostnike in starše Ljubljana.
- Kolkman, M. E., Kroesbergen, E. H., & Leseman, P. P. M. (2013). Early numerical development and the role of nonsymbolic and symbolic skills. *Learning and Instruction*, 25, 95–103. <https://doi.org/10.1016/j.learninstruc.2012.12.001>
- Landerl, K. (2013). Development of numerical processing in children with typical and dyscalculic arithmetic skills—A longitudinal study. *Frontiers in Psychology*, 4, Article 459. <https://doi.org/10.3389/fpsyg.2013.00459>
- Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8–9-year-old students. *Cognition*, 93(2), 99–125. <https://doi.org/10.1016/j.cognition.2003.11.004>
- Landerl, K., Fussenegger, B., Moll, K., & Willburger, E. (2009). Dyslexia and dyscalculia: Two learning disorders with different cognitive profiles. *Journal of Experimental Child Psychology*, 103(3), 309–324. <https://doi.org/10.1016/j.jecp.2009.03.006>
- Landerl, K., & Kölle, C. (2009). Typical and atypical development of basic numerical skills in elementary school. *Journal of Experimental Child Psychology*, 103(4), 546–565. <https://doi.org/10.1016/j.jecp.2008.12.006>
- Laski, E. V., & Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development*, 78(6), 1723–1743. <https://doi.org/10.1111/j.1467-8624.2007.01087.x>
- LeFevre, J.-A., Jimenez Lira, C., Sowinski, C., Cankaya, O., Kamawar, D., & Skwarchuk, S.-L. (2013). Charting the role of the number line in mathematical development. *Frontiers in Psychology*, 4, Article 641. <https://doi.org/10.3389/fpsyg.2013.00641>
- Lourenco, S. F., & Bonny, J. W. (2017). Representations of numerical and non-numerical magnitude both contribute to mathematical competence in children. *Developmental Science*, 20(4), Article e12418, 1–16. <https://doi.org/10.1111/desc.12418>
- Lourenco, S. F., Bonny, J. W., Fernandez, E. P., & Rao, S. (2012). Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence. *Proceedings of the National Academy of Sciences*, 109(46), 18737–18742. <https://doi.org/10.1073/pnas.1207212109>
- Magajna, L., Kavkler, M., Čačinovič Vogrinčič, G., Pečjak, S., & Bregar Golobič, K. (2008). *Učne težave v osnovni šoli: Koncept dela* [Learning difficulties in primary school: Work concept]. Zavod republike Slovenije za šolstvo.
- Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE*, 6(9), Article e23749. <https://doi.org/10.1371/journal.pone.0023749>
- McGraw-Hill Education (n.d.). *The Number Knowledge Test*.
- Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychological Science*, 24(10), 2013–2019. <https://doi.org/10.1177/0956797613482944>
- Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Science*, 14(12), 542–551. <https://doi.org/10.1016/j.tics.2010.09.008>
- Praet, M., & Desoete, A. (2014). Number line estimation from kindergarten to grade 2: A longitudinal study. *Learning and Instruction*, 33, 19–28. <https://doi.org/10.1016/j.learninstruc.2014.02.003>
- Rousselle, L., & Noël, M.-P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. *Cognition*, 102, 361–395. <https://doi.org/10.1016/j.cognition.2006.01.005>
- Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *British Journal of Developmental Psychology*, 30(2), 344–357. <https://doi.org/10.1111/j.2044-835X.2011.02048.x>
- Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number–space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology*, 114(3), 418–431. <https://doi.org/10.1016/j.jecp.2012.10.012>
- Schleepen, T. M. J., Van Mier, H. I., & De Smedt, B. (2016). The contribution of numerical magnitude comparison and phonological processing to individual differences in fourth graders' multiplication fact ability. *PLoS ONE*, 11(6), Article e0158335. <https://doi.org/10.1371/journal.pone.0158335>
- Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., & Luwel, K. (2018). Associations of number line estimation with mathematical competence: A meta-analysis. *Child Development*, 89(5), 1467–1484. <https://doi.org/10.1111/cdev.13068>
- Siegler, R. S. (2016). Magnitude knowledge: The common core of numerical development. *Developmental Science*, 19(3), 341–361. <https://doi.org/10.1111/desc.12395>

- Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, 75(2), 428–444. <https://doi.org/10.1111/j.1467-8624.2004.00684.x>
- Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, 14(3), 237–250. <https://doi.org/10.1111/1467-9280.02438>
- Wilson, A. J., Revkin, S. K., Cohen, D., Cohen, L., & Dehaene, S. (2006). An open trial assessment of “The Number Race”, an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions*, 2(1), Article 20. <https://doi.org/10.1186/1744-9081-2-20>
- Xenidou-Dervou, I., Molenaar, D., Ansari, D., van der Schoot, M., & van Lieshout, E. C. D. M. (2017). Nonsymbolic and symbolic magnitude comparison skills as longitudinal predictors of mathematical achievement. *Learning and Instruction*, 50, 1–13. <https://doi.org/10.1016/j.learninstruc.2016.11.001>
- Yang, D.-C., & Li, M. F. (2008). An investigation of 3rd-grade Taiwanese students’ performance in number sense. *Educational Studies*, 34(5), 443–455. <https://doi.org/10.1080/03055690802288494>
- Yang, D.-C., & Lin, Y.-C. (2015). Assessing 10- to 11-year-old children’s performance and misconceptions in number sense using a four-tier diagnostic test. *Educational Research*, 57(4), 368–388. <https://doi.org/10.1080/00131881.2015.1085235>

## Appendix

See Tables A1-A3.

*Prispelo/Received: 25. 8. 2021*  
*Sprejeto/Accepted: 7. 5. 2024*



**Table A1**  
*Psychometric Information for Items on Both Number Line Estimation Tasks, Separately for Each Grade*

Item	Item content	Grade	Absolute error (M and SD)	Item-total correlation (with 95% CI)	Cronbach's alpha if item deleted (with 95% CI)
Number line estimation – NP task					
Item 1	Mark position of number 3	1 <sup>st</sup>	11.73 (19.62)	.557 [.409, .677]	.758 [.681, .823]
		2 <sup>nd</sup>	5.39 (6.54)	.194 [.009, .365]	.810 [.751, .860]
		3 <sup>rd</sup>	3.47 (4.33)	.478 [.291, .630]	.647 [.519, .751]
Item 2	Mark position of number 47	1 <sup>st</sup>	16.43 (13.20)	.436 [.265, .580]	.775 [.703, .835]
		2 <sup>nd</sup>	11.27 (8.86)	.378 [.208, .526]	.796 [.733, .849]
		3 <sup>rd</sup>	8.82 (7.10)	.343 [.135, .522]	.655 [.531, .757]
Item 3	Mark position of number 92	1 <sup>st</sup>	20.17 (21.31)	.578 [.434, .693]	.755 [.677, .820]
		2 <sup>nd</sup>	11.48 (16.31)	.544 [.400, .661]	.783 [.715, .840]
		3 <sup>rd</sup>	6.14 (5.61)	.544 [.371, .681]	.627 [.492, .738]
Item 4	Mark position of number 25	1 <sup>st</sup>	24.84 (20.15)	.550 [.400, .672]	.759 [.682, .823]
		2 <sup>nd</sup>	10.67 (10.75)	.412 [.247, .554]	.795 [.731, .849]
		3 <sup>rd</sup>	6.20 (4.82)	.360 [.155, .535]	.657 [.533, .759]
Item 5	Mark position of number 64	1 <sup>st</sup>	13.64 (12.47)	.270 [.082, .440]	.790 [.723, .846]
		2 <sup>nd</sup>	8.40 (10.10)	.613 [.482, .717]	.770 [.698, .830]
		3 <sup>rd</sup>	7.24 (6.74)	.314 [.104, .497]	.661 [.538, .762]
Item 6	Mark position of number 18	1 <sup>st</sup>	24.88 (20.85)	.546 [.394, .668]	.760 [.683, .824]
		2 <sup>nd</sup>	10.62 (11.52)	.427 [.264, .566]	.793 [.728, .847]
		3 <sup>rd</sup>	7.65 (9.23)	.417 [.220, .582]	.643 [.514, .749]
Item 7	Mark position of number 86	1 <sup>st</sup>	19.11 (16.29)	.268 [.079, .439]	.792 [.727, .848]
		2 <sup>nd</sup>	15.25 (13.51)	.559 [.416, .674]	.772 [.702, .832]
		3 <sup>rd</sup>	12.66 (9.09)	.384 [.182, .555]	.651 [.525, .755]
Item 8	Mark position of number 33	1 <sup>st</sup>	20.64 (17.36)	.566 [.419, .685]	.758 [.681, .822]
		2 <sup>nd</sup>	10.82 (11.08)	.394 [.226, .540]	.794 [.730, .848]
		3 <sup>rd</sup>	8.29 (8.85)	.213 [-.004, .411]	.692 [.580, .783]
Item 9	Mark position of number 52	1 <sup>st</sup>	14.69 (11.92)	.486 [.324, .621]	.772 [.699, .833]
		2 <sup>nd</sup>	8.91 (9.46)	.649 [.528, .744]	.771 [.700, .831]
		3 <sup>rd</sup>	5.95 (6.57)	.384 [.182, .555]	.651 [.524, .754]

Table A1 (Continued)

Item	Item content	Grade	Absolute error (M and SD)	Item-total correlation (with 95% CI)	Cronbach's alpha if item deleted (with 95% CI)
Item 10	Mark position of number 71	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	13.74 (14.23) 8.24 (11.18) 5.22 (4.61)	.341 [.158, .500] .566 [.426, .680] .125 [-.095, .333]	.784 [.715, .842] .775 [.705, .834] .687 [.573, .780]
Number line estimation – PN task					
Cronbach's alpha = .813/.804/.725					
Item 1	Which number belongs to this position? [74]	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	29.44 (21.90) 13.23 (12.63) 10.90 (6.75)	.620 [.484, .726] .578 [.440, .689] .416 [.220, .580]	.771 [.688, .839] .774 [.704, .832] .699 [.590, .788]
Item 2	Which number belongs to this position? [39]	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	21.44 (11.13) 11.46 (8.89) 7.33 (5.13)	.517 [.357, .647] .440 [.277, .578] .477 [.292, .628]	.793 [.718, .854] .791 [.727, .845] .691 [.580, .782]
Item 3	Which number belongs to this position? [82]	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	26.19 (24.75) 11.23 (9.78) 8.38 (3.84)	.565 [.416, .685] .606 [.474, .711] .466 [.278, .619]	.789 [.714, .851] .771 [.700, .830] .697 [.588, .787]
Item 4	Which number belongs to this position? [56]	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	26.37 (15.53) 10.81 (10.79) 8.61 (8.94)	.682 [.562, .774] .489 [.335, .618] .624 [.472, .740]	.770 [.688, .838] .786 [.721, .842] .652 [.527, .755]
Item 5	Which number belongs to this position? [15]	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	11.13 (12.50) 9.03 (11.29) 5.45 (3.40)	.204 [.012, .382] .378 [.208, .526] .344 [.139, .521]	.822 [.757, .875] .802 [.741, .853] .713 [.609, .798]
Item 6	Which number belongs to this position? [6]	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	6.63 (11.97) 3.54 (3.81) 2.60 (2.25)	.395 [.219, .546] .482 [.327, .612] .278 [.067, .466]	.811 [.741, .867] .798 [.735, .850] .721 [.621, .803]
Item 7	Which number belongs to this position? [43]	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	23.39 (12.41) 11.21 (9.01) 7.61 (7.98)	.564 [.415, .684] .450 [.288, .586] .513 [.333, .657]	.791 [.716, .853] .789 [.724, .843] .680 [.565, .774]
Item 8	Which number belongs to this position? [98]	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	10.73 (25.31) 2.42 (9.33) 0.83 (1.61)	.464 [.290, .609] .565 [.425, .679] .217 [.001, .413]	.814 [.749, .867] .777 [.708, .835] .726 [.628, .807]

**Table A1** (Continued)

Item	Item content	Grade	Absolute error (M and SD)	Item-total correlation (with 95% CI)	Cronbach's alpha if item deleted (with 95% CI)
Item 9	Which number belongs to this position? [67]	1 <sup>st</sup>	28.42 (18.98)	.711 [.598, .796]	.768 [.683, .837]
		2 <sup>nd</sup>	11.45 (9.52)	.555 [.412, .671]	.778 [.710, .835]
		3 <sup>rd</sup>	10.13 (7.33)	.373 [.171, .545]	.709 [.605, .795]
Item 10	Which number belongs to this position? [20]	1 <sup>st</sup>	11.05 (8.22)	.313 [.124, .481]	.822 [.759, .875]
		2 <sup>nd</sup>	9.86 (11.37)	.401 [.232, .546]	.799 [.738, .851]
		3 <sup>rd</sup>	6.51 (6.35)	.287 [.076, .474]	.721 [.621, .804]

**Table A2**  
*Psychometric Information for Retained Items on the Area Comparison Task (with 95% Confidence Intervals in Brackets), Separately for Each Grade (Cronbach's Alpha = .796/.718/.683)*

Item no.	Correct answer	Position	Grade	Percent correct	Item-total correlation	Cronbach's alpha if item deleted
Ratio 6:5 (average percent correct = 82.9/85.8/85.1)						
Item 3	dark	right	1 <sup>st</sup>	91.5 [86.4, 96.6]	.183 [.002, .353]	.794 [.737, .844]
			2 <sup>nd</sup>	91.3 [86.1, 96.5]	.137 [-.048, .312]	.716 [.636, .785]
			3 <sup>rd</sup>	91.7 [85.8, 97.6]	-.038 [-.251, .178]	.687 [.584, .776]
Item 7	light	left	1 <sup>st</sup>	91.5 [86.4, 96.6]	.315 [.141, .469]	.791 [.734, .842]
			2 <sup>nd</sup>	93.9 [89.5, 98.3]	.370 [.201, .518]	.709 [.628, .780]
			3 <sup>rd</sup>	97.6 [94.3, 100.9]	.136 [-.081, .340]	.681 [.575, .771]
Item 20	dark	left	1 <sup>st</sup>	86.3 [80.1, 92.5]	.319 [.146, .473]	.791 [.733, .841]
			2 <sup>nd</sup>	92.2 [87.3, 97.1]	.067 [-.118, .247]	.718 [.639, .787]
			3 <sup>rd</sup>	89.3 [82.7, 95.9]	.228 [.014, .422]	.676 [.569, .768]
Item 24	light	right	1 <sup>st</sup>	77.8 [70.3, 85.3]	.142 [-.041, .315]	.796 [.739, .845]
			2 <sup>nd</sup>	77.4 [69.8, 85.0]	.134 [-.051, .309]	.716 [.637, .786]
			3 <sup>rd</sup>	78.6 [69.8, 87.4]	.212 [-.003, .408]	.676 [.569, .768]
Item 25	dark	left	1 <sup>st</sup>	81.2 [74.1, 88.3]	.312 [.139, .467]	.790 [.732, .841]
			2 <sup>nd</sup>	82.6 [75.7, 89.5]	.058 [-.126, .239]	.720 [.641, .788]
			3 <sup>rd</sup>	65.5 [55.3, 75.7]	.054 [-.162, .266]	.687 [.583, .775]
Item 28	dark	right	1 <sup>st</sup>	79.9 [72.6, 87.2]	.322 [.149, .475]	.790 [.732, .841]
			2 <sup>nd</sup>	75.7 [67.9, 83.5]	.297 [.120, .455]	.708 [.626, .780]
			3 <sup>rd</sup>	84.5 [76.8, 92.2]	.233 [.019, .426]	.675 [.568, .767]
Item 42	light	left	1 <sup>st</sup>	76.1 [68.4, 83.8]	.465 [.310, .596]	.785 [.726, .837]
			2 <sup>nd</sup>	84.3 [77.7, 90.9]	.273 [.095, .435]	.710 [.629, .781]
			3 <sup>rd</sup>	90.5 [84.2, 96.8]	.355 [.152, .529]	.671 [.562, .764]
Item 43	light	right	1 <sup>st</sup>	78.6 [71.2, 86.0]	.413 [.251, .553]	.787 [.728, .839]
			2 <sup>nd</sup>	88.7 [82.9, 94.5]	.227 [.045, .393]	.713 [.632, .783]
			3 <sup>rd</sup>	83.3 [75.3, 91.3]	.186 [-.029, .385]	.678 [.571, .769]
Ratio 7:6 (average percent correct = 79.5/83.0/86.9)						
Item 4	dark	left	1 <sup>st</sup>	92.3 [87.5, 97.1]	.422 [.260, .560]	.789 [.731, .841]
			2 <sup>nd</sup>	93.9 [89.5, 98.3]	.240 [.059, .405]	.713 [.633, .783]
			3 <sup>rd</sup>	97.6 [94.3, 100.9]	.170 [-.047, .370]	.680 [.574, .771]



Table A2 (Continued)

Item no.	Correct answer	Position	Grade	Percent correct	Item-total correlation	Cronbach's alpha if item deleted
Item 9	light	left	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	88.0 [82.1, 93.9] 87.8 [81.8, 93.8] 95.2 [90.6, 99.8]	-.065 [-.244, .118] .083 [-.102, .262] .148 [-.068, .351]	.800 [.745, .848] .718 [.639, .787] .680 [.574, .771]
Item 10	light	right	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	86.3 [80.1, 92.5] 87.8 [81.8, 93.8] 85.7 [78.2, 93.2]	.003 [-.179, .185] .252 [.072, .416] .016 [-.199, .229]	.799 [.743, .847] .711 [.631, .782] .686 [.582, .775]
Item 16	dark	right	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	70.1 [61.8, 78.4] 79.1 [71.7, 86.5] 81.0 [72.6, 89.4]	.319 [.146, .473] .005 [-.179, .187] .135 [-.081, .340]	.790 [.732, .841] .723 [.645, .790] .680 [.575, .771]
Item 33	light	left	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	88.9 [83.2, 94.6] 85.2 [78.7, 91.7] 82.1 [73.9, 90.3]	.339 [.168, .490] .103 [-.082, .281] .001 [-.213, .216]	.790 [.732, .841] .717 [.638, .787] .688 [.584, .776]
Item 34	dark	left	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	76.9 [69.3, 84.5] 81.7 [74.6, 88.8] 90.5 [84.2, 96.8]	.229 [.049, .394] .125 [-.060, .301] .238 [.025, .430]	.793 [.736, .843] .717 [.637, .786] .676 [.568, .768]
Item 38	light	right	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	77.8 [70.3, 85.3] 81.7 [74.6, 88.8] 89.3 [82.7, 95.9]	.377 [.209, .522] .297 [.120, .455] .253 [.041, .444]	.788 [.730, .840] .709 [.627, .780] .675 [.567, .767]
Item 45	dark	right	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	55.6 [46.6, 64.6] 67.0 [58.4, 75.6] 73.8 [64.4, 83.2]	.457 [.301, .590] .315 [.140, .471] .174 [-.042, .375]	.785 [.725, .837] .707 [.625, .778] .678 [.572, .769]
Ratio 8.7 (average percent correct = 74.2/77.2/77.2)						
Item 1	light	left	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	94.0 [89.7, 98.3] 85.2 [78.7, 91.7] 91.7 [85.8, 97.6]	.306 [.132, .462] .151 [-.033, .325] .147 [-.070, .350]	.792 [.734, .843] .715 [.636, .785] .680 [.574, .770]
Item 5	dark	left	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	68.4 [60.0, 76.8] 65.2 [56.5, 73.9] 64.3 [54.1, 74.5]	.132 [-.051, .306] .061 [-.124, .241] .307 [.099, .489]	.796 [.740, .846] .721 [.643, .789] .669 [.559, .763]
Item 17	dark	right	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	74.4 [66.5, 82.3] 80.0 [72.7, 87.3] 84.5 [76.8, 92.2]	.045 [-.137, .225] .184 [.000, .355] .103 [-.114, .310]	.799 [.743, .848] .714 [.634, .784] .682 [.577, .772]

Table A2 (Continued)

Item no.	Correct answer	Position	Grade	Percent correct	Item-total correlation	Cronbach's alpha if item deleted
Item 18	light	right	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	83.8 [77.1, 90.5] 86.1 [79.8, 92.4] 85.7 [78.2, 93.2]	.110 [-.073, .286] .403 [.237, .546] .299 [.090, .482]	.796 [.740, .846] .705 [.622, .777] .672 [.563, .765]
Item 30	dark	right	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	83.8 [77.1, 90.5] 88.7 [82.9, 94.5] 78.6 [69.8, 87.4]	.306 [.132, .462] .150 [-.034, .324] .199 [-.016, .397]	.791 [.733, .842] .715 [.636, .785] .677 [.570, .768]
Item 32	dark	left	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	63.2 [54.5, 71.9] 66.1 [57.4, 74.8] 65.5 [55.3, 75.7]	.156 [-.026, .328] .110 [-.074, .287] -.037 [-.250, .179]	.796 [.739, .845] .718 [.639, .787] .693 [.591, .780]
Item 47	light	left	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	55.6 [46.6, 64.6] 67.8 [59.3, 76.3] 72.6 [63.1, 82.1]	.392 [.226, .535] .239 [.058, .404] .175 [-.041, .375]	.787 [.728, .839] .711 [.630, .782] .678 [.572, .769]
Item 48	light	right	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	70.1 [61.8, 78.4] 78.3 [70.8, 85.8] 75.0 [65.7, 84.3]	.434 [.275, .571] .339 [.166, .491] .328 [.122, .507]	.786 [.727, .838] .706 [.624, .778] .668 [.558, .762]
Ratio 9:8 (average percent correct = 71.3/75.4/79.2)						
Item 2	light	right	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	88.9 [83.2, 94.6] 87.8 [81.8, 93.8] 94.0 [88.9, 99.1]	.194 [.013, .363] .231 [.050, .397] .102 [-.115, .310]	.794 [.737, .844] .712 [.632, .783] .682 [.576, .772]
Item 8	light	left	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	74.4 [66.5, 82.3] 76.5 [68.8, 84.2] 75.0 [65.7, 84.3]	.269 [.092, .430] .220 [.038, .387] .174 [-.042, .374]	.792 [.734, .842] .712 [.632, .783] .678 [.572, .769]
Item 12	dark	left	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	78.6 [71.2, 86.0] 83.5 [76.7, 90.3] 92.9 [87.4, 98.4]	.061 [-.122, .240] .128 [-.057, .304] .267 [.056, .455]	.798 [.742, .847] .716 [.637, .786] .675 [.568, .767]
Item 13	dark	right	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	86.3 [80.1, 92.5] 86.1 [79.8, 92.4] 94.0 [88.9, 99.1]	.319 [.146, .473] .124 [-.060, .301] .201 [-.014, .398]	.791 [.733, .841] .716 [.637, .786] .678 [.572, .769]
Item 31	light	right	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	79.5 [72.2, 86.8] 80.0 [72.7, 87.3] 72.6 [63.1, 82.1]	.240 [.061, .404] .333 [.160, .487] .140 [-.077, .344]	.793 [.735, .843] .707 [.625, .778] .681 [.575, .771]

Table A2 (Continued)

Item no.	Correct answer	Position	Grade	Percent correct	Item-total correlation	Cronbach's alpha if item deleted
Item 37	light	left	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	82.1 [75.2, 89.0] 84.3 [77.7, 90.9] 91.7 [85.8, 97.6]	.461 [.306, .593] .376 [.207, .523] .222 [.008, .417]	.786 [.727, .838] .706 [.623, .778] .677 [.570, .768]
Item 44	dark	right	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	42.7 [33.7, 51.7] 55.7 [46.6, 64.8] 52.4 [41.7, 63.1]	.199 [.019, .368] .203 [.021, .372] .054 [−.162, .265]	.794 [.737, .844] .713 [.633, .783] .687 [.584, .776]
Item 59	dark	dark	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	37.6 [28.8, 46.4] 49.6 [40.5, 58.7] 60.7 [50.3, 71.1]	.212 [.032, .379] .326 [.152, .480] .149 [−.067, .353]	.794 [.736, .844] .706 [.623, .778] .680 [.574, .771]
Ratio 10:9 (average percent correct = 63.0/69.1/73.2)						
Item 6	light	right	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	79.5 [72.2, 86.8] 77.4 [69.8, 85.0] 69.0 [59.1, 78.9]	.068 [−.115, .247] .327 [.153, .481] .301 [.093, .484]	.798 [.742, .847] .707 [.625, .778] .670 [.560, .763]
Item 19	light	left	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	85.5 [79.1, 91.9] 86.1 [79.8, 92.4] 88.1 [81.2, 95.0]	.146 [−.037, .319] .224 [.043, .391] .163 [−.053, .365]	.795 [.738, .845] .712 [.632, .783] .679 [.573, .770]
Item 22	dark	left	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	77.8 [70.3, 85.3] 80.9 [73.7, 88.1] 84.5 [76.8, 92.2]	.307 [.133, .463] .050 [−.135, .231] .291 [.082, .476]	.791 [.732, .841] .720 [.642, .789] .672 [.563, .765]
Item 29	light	right	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	69.2 [60.8, 77.6] 69.6 [61.2, 78.0] 66.7 [56.6, 76.8]	.196 [.015, .365] .083 [−.102, .262] .100 [−.117, .308]	.794 [.737, .844] .720 [.641, .788] .684 [.579, .773]
Item 36	light	left	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	59.8 [50.9, 68.7] 63.5 [54.7, 72.3] 67.9 [57.9, 77.9]	.034 [−.149, .214] .054 [−.131, .235] .262 [.050, .451]	.800 [.745, .849] .722 [.644, .790] .672 [.564, .765]
Item 41	dark	right	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	51.3 [42.2, 60.4] 53.0 [43.9, 62.1] 70.2 [60.4, 80.0]	.176 [−.006, .347] .225 [.043, .392] .120 [−.096, .326]	.795 [.738, .845] .712 [.631, .782] .682 [.577, .772]
Item 58	dark	dark	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	47.9 [38.8, 57.0] 67.8 [59.3, 76.3] 73.8 [64.4, 83.2]	.339 [.168, .490] .292 [.115, .451] .253 [.041, .443]	.789 [.731, .840] .708 [.626, .779] .673 [.565, .766]

Table A2 (Continued)

Item no.	Correct answer	Position	Grade	Percent correct	Item-total correlation	Cronbach's alpha if item deleted
Item 62	dark	left	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	33.3 [24.8, 41.8] 54.8 [45.7, 63.9] 65.5 [55.3, 75.7]	.096 [-.087, .273] .142 [-.042, .317] .276 [.066, .463]	.798 [.742, .847] .717 [.638, .786] .671 [.562, .764]
Ratio 11:10 (average percent correct = 64.0/70.4/73.8)						
Item 11	light	left	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	63.2 [54.5, 71.9] 60.0 [51.0, 69.0] 60.7 [50.3, 71.1]	.266 [.089, .427] .251 [.071, .415] .101 [-.116, .309]	.792 [.734, .842] .710 [.629, .781] .684 [.579, .773]
Item 14	dark	left	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	82.1 [75.2, 89.0] 85.2 [78.7, 91.7] 90.5 [84.2, 96.8]	.169 [-.013, .340] .084 [-.101, .263] .158 [-.058, .360]	.795 [.738, .844] .718 [.639, .787] .679 [.573, .770]
Item 23	light	right	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	83.3 [76.5, 90.1] 84.3 [77.7, 90.9] 81.0 [72.6, 89.4]	.236 [.057, .400] .003 [-.180, .186] .169 [-.048, .370]	.793 [.735, .843] .722 [.644, .790] .679 [.572, .769]
Item 26	dark	left	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	59.8 [50.9, 68.7] 63.5 [54.7, 72.3] 66.7 [56.6, 76.8]	.330 [.158, .483] .097 [-.088, .275] .002 [-.213, .216]	.789 [.731, .840] .719 [.641, .788] .690 [.588, .778]
Item 35	light	right	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	64.1 [55.4, 72.8] 71.3 [63.0, 79.6] 78.6 [69.8, 87.4]	.245 [.066, .408] .173 [-.011, .345] .336 [.131, .513]	.793 [.735, .843] .715 [.635, .784] .668 [.558, .762]
Item 39	light	left	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	65.8 [57.2, 74.4] 76.5 [68.8, 84.2] 73.8 [64.4, 83.2]	.358 [.189, .506] .295 [.118, .453] .234 [.021, .427]	.788 [.730, .840] .708 [.627, .780] .674 [.566, .766]
Item 40	dark	right	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	52.1 [43.0, 61.2] 65.2 [56.5, 73.9] 73.8 [64.4, 83.2]	.404 [.240, .545] .255 [.075, .419] .180 [-.036, .380]	.787 [.727, .838] .710 [.629, .781] .678 [.571, .769]
Item 57	dark	dark	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	41.9 [33.0, 50.8] 57.4 [48.4, 66.4] 65.5 [55.3, 75.7]	.151 [-.032, .323] .185 [.001, .356] .248 [.035, .439]	.796 [.739, .845] .714 [.634, .784] .673 [.565, .766]



**Table A3**

*Psychometric Information for Items on the Number Comparison Test (with 95% Confidence Intervals in Brackets), Separately for Each Grade (Average Percent Correct = 59.9/67.0/76.0; Cronbach's Alpha = .628/.624/.770)*

Item no.	Item detail	Grade	Percent correct	Item-total correlation	Cronbach's alpha if item deleted
Item 1	Triangle 3-1-7	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	79.5 [72.2, 86.8] 89.6 [84.0, 95.2] 94.0 [88.9, 99.1]	.387 [.218, .533] .418 [.254, .559] .411 [.215, .574]	.601 [.486, .699] .597 [.482, .697] .760 [.679, .828]
Item 2	Triangle 7-5-10	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	63.2 [54.5, 71.9] 82.6 [75.7, 89.5] 85.7 [78.2, 93.2]	.224 [.042, .392] .172 [−.013, .345] .413 [.218, .576]	.615 [.505, .710] .617 [.505, .710] .757 [.675, .826]
Item 3	Triangle 6-3-8	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	77.8 [70.3, 85.3] 83.5 [76.7, 90.3] 91.7 [85.8, 97.6]	.245 [.064, .410] .435 [.273, .573] .224 [.010, .418]	.614 [.503, .709] .590 [.473, .692] .767 [.688, .833]
Item 4	Triangle 5-8-1	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	66.7 [58.2, 75.2] 69.6 [61.2, 78.0] 72.6 [63.1, 82.1]	.099 [−.086, .278] .240 [.058, .406] .327 [.121, .506]	.629 [.523, .721] .610 [.498, .706] .761 [.681, .829]
Item 5	Triangle 4-9-1	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	76.1 [68.4, 83.8] 82.6 [75.7, 89.5] 90.5 [84.2, 96.8]	.322 [.147, .478] .372 [.202, .521] .366 [.164, .538]	.612 [.501, .708] .617 [.507, .711] .752 [.668, .823]
Item 6	Triangle 8-10-4	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	77.8 [70.3, 85.3] 87.8 [81.8, 93.8] 86.9 [79.7, 94.1]	.264 [.084, .427] .182 [−.002, .354] .508 [.330, .651]	.606 [.493, .703] .596 [.481, .696] .760 [.680, .829]
Item 7	Triangle 13-18-10	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	67.5 [59.0, 76.0] 86.1 [79.8, 92.4] 94.0 [88.9, 99.1]	.277 [.098, .438] .373 [.203, .521] .338 [.134, .515]	.609 [.498, .706] .598 [.484, .698] .763 [.684, .848]
Item 8	Triangle 18-20-14	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	60.7 [51.8, 69.6] 84.3 [77.7, 90.9] 86.9 [79.7, 94.1]	.228 [.046, .396] .337 [.164, .491] .519 [.342, .660]	.615 [.505, .710] .601 [.487, .700] .751 [.668, .822]
Item 9	Triangle 17-13-19	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	70.9 [62.7, 79.1] 88.7 [82.9, 94.5] 91.7 [85.8, 97.6]	.393 [.225, .538] .377 [.208, .525] .285 [.075, .471]	.597 [.482, .697] .600 [.486, .699] .764 [.685, .831]
Item 10	Triangle 14-12-17	1 <sup>st</sup> 2 <sup>nd</sup> 3 <sup>rd</sup>	68.4 [60.0, 76.8] 84.3 [77.7, 90.9] 97.4 [94.0, 100.8]	.424 [.260, .564] .346 [.173, .498] .374 [.173, .544]	.593 [.476, .693] .600 [.486, .699] .760 [.679, .828]

Table A3 (Continued)

Item no.	Item detail	Grade	Percent correct	Item-total correlation	Cronbach's alpha if item deleted
Item 11	Triangle 16-18-13	1 <sup>st</sup>	59.0 [50.1, 67.9]	.254 [.073, .418]	.612 [.501, .708]
		2 <sup>nd</sup>	76.5 [68.8, 84.2]	.326 [.151, .481]	.600 [.501, .708]
		3 <sup>rd</sup>	88.1 [81.2, 95.0]	.437 [.245, .595]	.756 [.674, .826]
Item 12	Triangle 15-10-18	1 <sup>st</sup>	55.6 [46.6, 64.6]	.210 [.027, .379]	.617 [.507, .712]
		2 <sup>nd</sup>	65.2 [56.5, 73.9]	.272 [.092, .434]	.605 [.492, .703]
		3 <sup>rd</sup>	83.3 [75.3, 91.3]	.312 [.105, .493]	.762 [.682, .830]
Item 13	Triangle 7-3-13	1 <sup>st</sup>	68.4 [60.0, 76.8]	.152 [-.033, .327]	.623 [.515, .716]
		2 <sup>nd</sup>	68.7 [60.2, 77.2]	.165 [-.019, .339]	.619 [.510, .713]
		3 <sup>rd</sup>	70.2 [60.4, 80.0]	.126 [-.091, .331]	.775 [.700, .839]
Item 14	Triangle 12-8-17	1 <sup>st</sup>	47.9 [38.8, 57.0]	.091 [-.095, .270]	.631 [.525, .722]
		2 <sup>nd</sup>	51.3 [42.2, 60.4]	.184 [.000, .356]	.617 [.508, .712]
		3 <sup>rd</sup>	65.5 [55.3, 75.7]	.282 [.072, .468]	.765 [.686, .832]
Item 15	Triangle 10-5-16	1 <sup>st</sup>	59.8 [50.9, 68.7]	.279 [.100, .440]	.609 [.497, .705]
		2 <sup>nd</sup>	70.4 [62.1, 78.7]	.138 [-.047, .314]	.622 [.514, .716]
		3 <sup>rd</sup>	86.9 [79.7, 94.1]	.269 [.058, .457]	.765 [.685, .832]
Item 16	Triangle 15-19-9	1 <sup>st</sup>	59.0 [50.1, 67.9]	.039 [-.146, .221]	.636 [.533, .726]
		2 <sup>nd</sup>	66.1 [57.4, 74.8]	-.047 [-.229, .138]	.646 [.545, .733]
		3 <sup>rd</sup>	70.2 [60.4, 80.0]	.133 [-.084, .338]	.775 [.699, .839]
Item 17	Triangle 9-13-7	1 <sup>st</sup>	70.9 [62.7, 79.1]	.291 [.113, .450]	.608 [.496, .705]
		2 <sup>nd</sup>	86.1 [79.8, 92.4]	.137 [-.048, .313]	.621 [.512, .714]
		3 <sup>rd</sup>	90.5 [84.2, 96.8]	.112 [-.105, .319]	.772 [.695, .837]
Item 18	Triangle 11-16-8	1 <sup>st</sup>	56.4 [47.4, 65.4]	.218 [.035, .386]	.616 [.506, .711]
		2 <sup>nd</sup>	71.3 [63.0, 79.6]	.274 [.094, .436]	.605 [.493, .703]
		3 <sup>rd</sup>	79.8 [71.2, 88.4]	.454 [.265, .609]	.753 [.670, .824]
Item 19	Triangle 21-25-18	1 <sup>st</sup>	32.5 [24.0, 41.0]	.154 [-.031, .329]	.623 [.515, .716]
		2 <sup>nd</sup>	28.7 [20.4, 37.0]	.148 [-.037, .323]	.621 [.513, .715]
		3 <sup>rd</sup>	36.9 [26.6, 47.2]	.383 [.184, .552]	.757 [.676, .827]
Item 20	Triangle 28-31-24	1 <sup>st</sup>	37.6 [28.8, 46.4]	.075 [-.110, .256]	.632 [.527, .723]
		2 <sup>nd</sup>	30.4 [22.0, 38.8]	.000 [-.184, .184]	.639 [.536, .729]
		3 <sup>rd</sup>	35.7 [25.5, 45.9]	.435 [.243, .594]	.753 [.670, .824]

**Table A3** (Continued)

Item no.	Item detail	Grade	Percent correct	Item-total correlation	Cronbach's alpha if item deleted
Item 21	Triangle 91-88-95	1 <sup>st</sup>	45.3 [36.3, 54.3]	-.060 [-.242, .125]	.648 [.547, .735]
		2 <sup>nd</sup>	13.9 [7.6, 20.2]	-.018 [-.201, .166]	.635 [.531, .725]
		3 <sup>rd</sup>	32.1 [22.1, 42.1]	.183 [-.033, .382]	.772 [.695, .837]
Item 22	Triangle 66-60-70	1 <sup>st</sup>	32.5 [24.0, 41.0]	.116 [-.069, .294]	.627 [.521, .719]
		2 <sup>nd</sup>	31.3 [22.8, 39.8]	-.015 [-.199, .169]	.641 [.539, .730]
		3 <sup>rd</sup>	52.4 [41.7, 63.1]	.310 [.102, .492]	.763 [.683, .831]
Item 23	Triangle 72-69-76	1 <sup>st</sup>	35.0 [26.4, 43.6]	.151 [-.034, .325]	.624 [.516, .717]
		2 <sup>nd</sup>	27.8 [19.6, 36.0]	.134 [-.051, .310]	.623 [.515, .716]
		3 <sup>rd</sup>	51.2 [40.5, 61.9]	.390 [.192, .558]	.757 [.675, .826]
Item 24	Triangle 50-47-52	1 <sup>st</sup>	69.2 [60.8, 77.6]	.157 [-.028, .331]	.623 [.515, .716]
		2 <sup>nd</sup>	80.9 [73.7, 88.1]	.196 [.013, .367]	.615 [.505, .710]
		3 <sup>rd</sup>	89.3 [82.7, 95.9]	.025 [-.191, .238]	.776 [.701, .840]

*Note:* Triangles are described as x-y-z, with x standing for the number in the upper corner (i.e., the reference number) and y and z standing for the numbers in the lower left corner and in the lower right corner, respectively (i.e., the comparison numbers).