Patterns of symbolic numerical magnitude processing and working memory as predictors of early mathematics performance

Riikka Mononen¹ · Markku Niemivirta^{2,3}

Received: 17 March 2021 / Revised: 9 November 2021 / Accepted: 21 December 2021 /
Published online: 4 February 2022
© The Author(s) 2022

Abstract

Although the roles of symbolic numerical magnitude processing (SNMP) and working memory (WM) in mathematics performance are well acknowledged, studies examining their joint effects are few. Here, we investigated the profiles of SNMP (1- and 2-digit comparison) and WM (verbal, visual and central executive) among Norwegian first graders ($N=256$), and how these predict performance in counting, arithmetic facts and word problem-solving. Using latent class cluster analysis, four groups were identified: (1) weak SNMP (33.6%), (2) strong SNMP (25.8%), (3) weak SNMP and WM (23.4%) and (4) strong WM (17.2%). Group differences in mathematics performance were significant with explained variance ranging from 7 to 16%, even after controlling for relevant demographics and domain-general cognitive skills. Our findings suggest that children may display relative strengths in SNMP and WM, and that they both have a unique, even compensatory role in mathematics performance.

Keywords Arithmetic · Counting · Latent class cluster analysis · Symbolic numerical magnitude processing · Working memory

Introduction

Several cross-sectional and longitudinal studies have looked at numerical magnitude processing (NMP) and working memory (WM) as possible domain-specific and domain-general cognitive precursors of children's mathematics performance and development (e.g. Alloway & Alloway, 2010; De Smedt et al., 2013; Friso-van den Bos et al., 2013; Kroesbergen & van Dijk, 2015; Li et al., 2018; Toll et al., 2016). Both NMP and WM have been

✉ Riikka Mononen
r.m.mononen@isp.uio.no

¹ Department of Special Needs Education, University of Oslo, Blindern, P.O. Box 1140, 0318 Oslo, Norway

² School of Applied Educational Science and Teacher Education, University of Eastern Finland, Joensuu, Finland

³ Department of Education, University of Helsinki, Helsinki, Finland

linked with individual differences in mathematics performance, and although weakness in either of them seem to result in inferior performance (e.g. Brankaer et al., 2017; Cañizares et al., 2012), surprisingly few studies have investigated them simultaneously (Chan & Wong, 2019; Kroesbergen & van Dijk, 2015; Passolunghi et al., 2014; Toll et al., 2016). The intriguing findings from these few studies, varying in designs and methods, on one hand suggest WM to be a stronger predictor than NMP of early mathematics performance (Passolunghi et al., 2014), but on the other hand, children having weakness in both NMP and WM to struggle most in mathematics compared to those having no weakness or weakness in either NMP or WM (Kroesbergen & van Dijk, 2015; Toll et al., 2016).

Most of these studies looking at how the patterns of NMP and WM among children predict mathematics performance have used measures of non-symbolic NMP or employed varying operationalisations of NMP, while recent research suggests that symbolic NMP (SNMP) might provide more consistent findings and stronger predictions on mathematics performance among school-aged children (De Smedt et al., 2013). Prior studies have often used visual WM as a measure of WM, but it would seem beneficial to include verbal WM and central executive in these investigations as well, as both are known to be associated with mathematics performance (Friso-van den Bos et al., 2013). Along with NMP and WM, several domain-general factors, such as fluid intelligence, rapid automatized naming and language skills, have also been linked with mathematics performance (Koponen et al., 2018; Purpura & Ganley, 2014), due to which further attention to these individual differences might be informative. Finally, as to the criterion variables, focus in terms of mathematics performance has mainly been on arithmetic facts and word problem-solving, while also some other core skills, such as counting, have shown to be important predictors of later mathematics performance (Aunola et al., 2004). Of particular interest is also the fact that previous studies have typically used *a priori* defined cut-off scores for defining different performance levels, and, consequently, grouping children accordingly to those levels (Kroesbergen & van Dijk, 2015; Toll et al., 2016). It would therefore seem informative to examine the extent to which an alternative, data-driven classification matched the previous findings.

With this study, we aim to expand on previous research on the joint role NMP and WM play in early mathematics performance. We will focus particularly on SNMP (1- and 2-digit comparison), due to the more robust and stronger predictions obtained with it, and include all three different components of WM for a more inclusive representation of WM. As to classifying children according to their performance profiles, we will extract empirical profiles of SNMP and WM by means of latent class clustering. Finally, we will also use a relatively comprehensive set of measures for mathematics performance and take into account the influence of various domain-general skills, in order to tease out better the unique effects of SNMP and WP on mathematics performance.

The relation between symbolic numerical magnitude processing and mathematics performance

The ability to approximately represent numerical magnitude, often called approximate number system (ANS), is considered to underlie both non-symbolic and symbolic numerical magnitude processing (Feigenson et al., 2013; Goffin & Ansari, 2019). The association between NMP and mathematics performance has been substantially investigated in recent years, and this association seems to partly depend on the number format used in the NMP tasks. Studies using a symbolic format (i.e. Arabic digits) have demonstrated more

consistent findings and stronger effects on mathematics performance than studies using non-symbolic format (i.e. dots) (De Smedt et al., 2013; Schneider et al., 2017). The inconsistencies regarding the non-symbolic format might partly be due to methodological issues, such as using many different types of non-symbolic comparison measures, or the measures not being sensitive enough. Alternatively, the non-symbolic magnitude processes measured may simply be less critical in the context of school mathematics (De Smedt et al., 2013). Given that these constraints do not seem to apply when NMP is measured using a symbolic format (De Smedt et al., 2013), we focused only on SNMP in this study.

SNMP predicts mathematics achievement within and across different grades in elementary school (Brankaer et al., 2017; Holloway & Ansari, 2009), also after controlling for age, intellectual ability and speed of number identification (De Smedt et al., 2009). Furthermore, students performing well in SNMP exhibit more effective arithmetic strategy use (i.e. being faster in retrieving facts and using procedural strategies), even when taking into account differences in intellectual ability, digit naming and general mathematics achievement (Vanbinst et al., 2012).

Tasks measuring SNMP aim to tap a person's ability to access the number magnitudes in symbols (Rousselle & Noël, 2007). A vast majority of studies have used 1-digit numbers on symbolic comparison tasks (Brankaer et al., 2017). Some studies show that children solve multi-digit comparison tasks slower than single-digit tasks (Landerl et al., 2009), which may stem from the fact that in such tasks, children process multi-digit numbers differently than single-digit numbers. Two alternative explanations have been proposed; one suggesting that children process the number as a uniform unit (i.e. holistic view) (Reynvoet & Brysbaert, 1999), and another suggesting that children process decade-digit and unit-digit of the number independently (i.e. compositional model) (Nuerk et al., 2001). In support of the compositional model, research has found children to compare compatible number pairs (i.e. when both digits of the number are bigger than in the number to be compared; 25 vs. 68) faster than incompatible number pairs (e.g. 51 vs. 37).

The role of WM in mathematics performance

A significant relation between WM and mathematics performance has been evidenced in several studies (Friso-van den Bos et al., 2013; for a meta-analysis). Previous research has often employed the multicomponent WM model by Baddeley and Hitch (1974), hence referring to the three subcomponents of WM: the two slave systems, phonological loop and visuospatial sketchpad for storing verbal and visuospatial information, respectively, and central executive for coordinating information of the slave systems. This three-component model is still drawing the most attention in educational research, even though a fourth component, episodic buffer, has been included in the model (Baddeley, 2010).

According to the meta-analysis by Friso-van den Bos et al. (2013), all three WM components are linked with children's mathematics performance. This relation, however, was dependent on the type of mathematics test used. General mathematics tests, such as national curriculum tests, which demand more switching between different operations and updating sets of information, yielded stronger correlations with WM than those focusing only on some specific mathematical skills. When looking at the relation between WM and specific mathematical skills, both verbal and visuospatial WM seem to be important predictors of counting (e.g. tasks involving number sequences or linking quantities to number words) at kindergarten age (Preßler et al., 2013). Kytälä et al. (2019) showed that verbal, but not visuospatial WM, predicted word problem-solving from kindergarten to second

grade, whereas Andersson (2008) found both verbal and central executive functions to predict word problem-solving among second to fourth graders. The use of central executive resources in solving single-digit arithmetic problems has also been highlighted, and verbal WM seems to play a role if such problems are solved using counting strategies (DeStefano & LeFevre, 2004, for a review).

However, the effect of WM seems to diminish when other factors are controlled. For example, basic academic skills (e.g. reading and calculation) and fluid intelligence have been found to account for some of the effects of WM on word problem-solving accuracy in early grades (Fung & Swanson, 2017; Zheng et al., 2011). Also, the effect of verbal WM on counting skills diminished when vocabulary, morphology, phonology, intelligence, task orientation and gender were controlled (Koponen et al., 2018).

SNMP and WM as predictors of mathematical learning difficulties

Mathematical learning difficulties (MLD) and dyscalculia (i.e. severe and persistent learning difficulties in mathematics) are another relevant issue in the present context. Given the impact of SNMP and WM on mathematics performance and development, children with MLD or dyscalculia would then be expected to display inferior SNMP and/or WM skills compared to their peers without such difficulties. Indeed, over the years, different theoretical models have been proposed as a cause of MLD (for an overview, see e.g. Siemann & Petermann, 2018), highlighting a deficit in domain-specific skills, namely in NMP—either non-symbolic NMP (i.e. defective number module, Butterworth, 2005) or SNMP (i.e. access deficit, Rousselle & Noël, 2007), or in domain-general skills, such as WM (i.e. cognitive deficit, Geary, 2004; Karagiannakis et al., 2014) or in both (i.e. double deficit, Kroesbergen & van Dijk, 2015; Wolf & Bowers, 1999). Regarding school beginners, as is the case in the current study, a weak performance in either NMP or WM or both might thus point out to a risk for MLD.

Some studies using non-symbolic NMP measures have shown children with dyscalculia to perform poorly in tasks of comparing two magnitudes (Desoete et al., 2012; Landerl et al., 2009; Mazzocco et al., 2011), or matching them (i.e. deciding whether two magnitudes are same or not) (Lafay et al., 2019), thus supporting the “defective number module” (Butterworth, 2005) as a primary cause of dyscalculia. Other studies, however, have not found such effects (e.g. De Smedt & Gilmore, 2011; Lafay et al., 2019). In contrast, more consistent and robust differences between children with or without MLD or dyscalculia (Cañizares et al., 2012; De Smedt et al., 2013; De Smedt & Gilmore, 2011; Desoete et al., 2012; Landerl et al., 2009) have been found in studies using SNMP measures, thus supporting “access deficit” (Rousselle & Noël, 2007) as an alternative explanation for dyscalculia, according to which the core deficit lies in not being able to access the number magnitude in symbols.

In support of the cognitive deficit as a cause for MLD, not only has it been shown that children with higher WM capacity outperform their peers with lower WM capacity in different mathematics tasks (e.g. Preßler et al., 2013), but also that children with varying degrees of MLD display weaker WM skills compared to their peers without MLD (e.g. Geary et al., 2004; Menon, 2016; Passolunghi & Siegel, 2004). In a meta-analysis comparing children with MLD to average-achieving age-matched children on measures of WM, large effects were found for central executive and visuospatial sketchpad ($d=0.95$ and $d=0.59$, respectively) and medium effects for phonological loop ($d=0.36$), in favour of the average-achieving children (David, 2012).

Investigating the roles of (S)NMP and WM simultaneously in mathematics performance

Surprisingly few studies have investigated the roles of NMP or SNMP and WM in mathematics performance, or in relation to MLD, simultaneously. Passolunghi et al. (2014) found WM to be a stronger predictor of mathematics performance than non-symbolic NMP in the beginning of the first grade (6-year-olds), after controlling for intelligence (i.e. verbal and fluid intelligence). Furthermore, NMP lost its significance by the end of the first grade when predicting teacher-rated mathematics performance in a similar manner. Note, however, that intelligence turned out to be an even stronger precursor of mathematics performance than either WM or NMP.

Recently, Chan and Wong (2019) tested whether the prediction of visuospatial WM at grade 1 (7-year-olds) on mathematics achievement at Grade 2 was mediated by numerical magnitude representation (i.e. SNMP and computation) and problem representation (i.e. word problems). Both pathways were found to be significant, even after controlling nonverbal intelligence, reading fluency, processing speed, and verbal WM. However, as the direct effect from visuospatial WM to mathematics achievement remained significant as well, the authors concluded that numerical magnitude representation and problem representation failed to fully explain the relation between visuospatial WM and mathematics achievement.

Kroesbergen and van Dijk (2015) examined visuospatial WM and NMP (although they used the term “number sense”, which was represented by a combined score of non-symbolic and symbolic comparison, and number line tasks) in relation to arithmetic fluency and word problem-solving among 6–10 years old, also including a subgroup of children with MLD. Using a cut-off point of scoring below the 25th percentile, the participants were first divided into four groups according to their performance in number sense and WM: weakness in number sense, weakness in WM, weakness in both number sense and WM (i.e. double weakness) and without weakness. When comparing these groups on mathematics performance, they found a deficit in either number sense or WM, or both, to be connected with lower performance. Those identified as having double weakness displayed the most inferior performance of all groups, even after controlling for age, IQ and verbal WM, thus supporting the double-deficit hypothesis of MLD.

Partly replicating the study by Kroesbergen and van Dijk (2015), Toll et al. (2016) also investigated the predictions of NMP (or “number sense”, in their study) and visuospatial WM from the first year of kindergarten (5-year-olds) to grade one (7-year-olds). Non-symbolic NMP was operationalised in terms of dot comparison tasks, and although conceptualised as “symbolic number sense”, which refers to SNMP in some other studies, the authors used counting-based tasks instead of a digit comparison task. The results showed non-symbolic NMP to predict only word problem-solving 2 years later, whereas symbolic number sense predicted both arithmetic facts and word problems. When grouping the participants scoring below the 25th percentiles, weakness in both visual WM and number sense was connected with the lowest performance in arithmetic facts and word problems, even when controlling for fluid intelligence, thus replicating the results of Kroesbergen and van Dijk (2015).

Current study

By expanding on previous captivating findings on the roles of SNMP and WM in mathematics performance, we will in this study first explore the patterning of SNMP and WM

through children's performance profiles, and then link these profiles with several aspects of children's mathematics performance (i.e. counting, arithmetic facts and word problem-solving). In contrast to previous studies using a priori defined cut-off points in performance (i.e. below or above the 25th percentile) to classify children into different groups, we examined the relative strength in children's SNPM and WM skills by extracting empirical profiles through latent class clustering. Thus, instead of forming fixed categories of specific combinations (e.g. high/low, high/high), we relied on more naturally occurring data-driven patterns representing groups of children similar to each other, but different from the others. In a sense, then, this approach might result in a somewhat less artificial account of the patterning of SNPM and WM among children. Prior studies have mainly used visuospatial WM as an indicator of WM, whereas we included three WM components—verbal, visual and central executive—in the classification, thus seeking to have a more comprehensive empirical representation of individual differences in WM. As to the outcome measures, we included verbal counting skills in addition to arithmetic facts (addition and subtraction) and word problem-solving used in previous studies (Kroesbergen & van Dijk, 2015; Toll et al., 2016), as it is one of the core skills developing in this age group, and a significant predictor of later mathematics performance (Aunola et al., 2004). Finally, we controlled for several demographics (i.e. age, gender, parental educational level, status of second language learner) and domain-general cognitive skills (i.e. fluid intelligence, rapid automatized naming and word comprehension) shown to be associated with children's mathematics performance (e.g. Alloway & Alloway, 2010; Koponen et al., 2018; Purpura & Ganley, 2014), to capture the unique contribution of children's skill profiles on mathematics performance.

Method

Participants

The study is part of a project, which traces individual differences in children's development of mathematical skills. The participants were 256 Norwegian first graders ($M_{age} = 6$ y. 9 m., $SD = 3.33$ m., 46.1% girls), from 12 classrooms in five public schools in the capital region of Norway (see demographics of participants in Table 1). Most of the schools were located nearby the data collection site, to avoid long transportation times for the children during the school day. Regarding the family socio-economic status, we used parents' highest educational level as a proxy, reported by parents in a questionnaire. These data indicate that our sample of children came mainly from families having middle or high levels of education. As reported by their parents, 10.6% of the children had Norwegian as their second language.

Before the data collection, an ethical approval was applied for and granted by the Norwegian Centre for Research Data, and consents for the participation from children's parents and teachers were collected accordingly.

Measures

Symbolic numerical magnitude processing

The SYmbolic Magnitude Processing (SYMP) Test (Brankaer et al., 2017) was used to measure children's SNPM. The SYMP consists of two subtests to be used with elementary

Table 1 Demographics of participants ($N=256$)

Variable	n (%)	M (SD)
Gender		
Girl	118 (46.1)	
Boy	138 (53.9)	
Age in months ^a		80.63 (3.33)
Norwegian as second language ^b	25 (10.9)	
Mother's highest educational level ^c		3.54 (0.82)
(1) Comprehensive school	3 (1.3)	
(2) Upper secondary school	22 (9.6)	
(3) Bachelor	70 (30.4)	
(4) Master	117 (50.9)	
(5) PhD	18 (7.8)	
Father's highest educational level ^d		3.51 (0.88)
(1) Comprehensive school	4 (1.7)	
(2) Upper secondary school	30 (13.1)	
(3) Bachelor	59 (25.8)	
(4) Master	118 (51.5)	
(5) PhD	18 (7.9)	

Note. ^{a-c} $n=230$, ^d $n=229$

school-aged children: a subtest with digits between 1 and 9 (SYMP 1-digit) and a subtest with digits ranging from 11 to 99 (SYMP 2-digit). Each subtest comprises 60 digit pairs, presented in four columns of 15 pairs. During the test, the child is asked to cross out the larger of the two digits. For both subtests, the children are given 30 s to solve as many items as possible. To ensure that the child understands the task, four practice trials are included in both subtests (for more details on item description and test validation, see Brankaer et al., 2017). One point is given for a correct answer and zero for an incorrect answer, the sum score for each subtest thus being the number of the items correctly solved in 30 s.

Working memory

WM was measured using three subtests of the standardised Wechsler Intelligence Scale for Children ages 6 through 16 (V Norwegian version) (Wechsler, 2017) to tap the different components of WM; *Forwards Digit Span* for the phonological loop, *Backwards Digit Span* for the central executive and *Picture Span* for the visual WM. In both Digit Span subtests, there are nine blocks in the test, each consisting of two items. The child needs to recall orally given digits (2–8) either forwards or backwards. The test is stopped if both items in the block are incorrect. One point is given for a correct answer and zero for an incorrect answer. In the Picture Span, the child is shown 2–8 pictures (such as a ball, flower and car) in an increasing number of pictures, for 5 s in the stimulus book. After this, the child needs to point these pictures in the same order from another collection of pictures. There are 25 items in the test (with 2 practice items). One point is given if the child is able to recall the correct pictures and two points if the order of them is also correct, thus the

total maximum score being 46 points. The test is stopped after three consecutive errors, as instructed in the handbook.

Mathematics outcome measures

Three different mathematical skills were measured: verbal counting and arithmetic facts (addition and subtraction), as well as word problem-solving. To measure children's verbal counting skills, we used the Number Sequences task, which is a subtest from a standardised LukiMat Mathematics test battery for first graders (Salminen & Koponen, 2011). This test measures children's verbal counting skills as number sequences (counting by 1 s, 2 s, 5 s and 10 s forwards and backwards) within the number range 1–100. There are 29 items in the task (13 counting forwards and 19 counting backwards), and the child responds orally to all tasks (i.e. reciting number sequences). One point is given for a correct answer and zero for an incorrect answer.

Addition and subtraction facts were measured using a standardised test Regnefaktaprøven [Test of arithmetic facts] (Klausen & Reikerås, 2016), designed for grades 2–10. We chose this measure to be used with first graders as the data collection was conducted at the end of the first grade, we needed the same measure to be used in different grades throughout the project years, and it had been developed and standardised in Norway. The test includes 45 items per page, either addition or subtraction problems, and 2 min are given for each subtest to solve as many problems as possible. One point is given for a correct answer and zero for an incorrect answer, the sum score for each subtest thus being the number of problems correctly solved in 2 min.

Word problem-solving was measured using the standardised test WISC-V: Regning [Arithmetic] (Wechsler, 2017), to be used with 6–16-year-olds. The test comprises 34 items. An arithmetic word problem is read aloud to the child one at a time. For the first five items, visual materials (i.e. pictures) are provided to support solving the problem, and 30 s are given to solve each problem. The test is stopped after three consecutive errors, as guided in the manual. One point is given for a correct answer and zero for an incorrect answer, or if the problem was not solved in time.

Covariates

To take into account additional relevant individual differences, a set of cognitive skills shown to affect mathematics performance was used as covariates in the analyses. *Fluid intelligence* was measured using a standardised test of Raven's Coloured Progressive Matrices (Raven et al., 1990), to be used with 5–11-year-old children. The test consists of 36 items in 3 sets (A, Ab, B), with 12 items per set. The child needs to find a piece out of six given possibilities that fits into the picture pattern. One point is given for a correct answer and zero for an incorrect answer. As the first two items in set A are practice items, the maximum total score thus is 34 points. *Rapid automatized naming* was measured using a subtest, Hurtig Benevning [Rapid naming] from the standardised test battery Clinical Evaluation of Language Fundamentals – Fourth Edition (CELF-4) (Semel et al., 2003) designed for 5–12-year-old children. This subtest consisted of two tasks, naming of colours and figures (i.e. non-alphanumeric RAN). The child needs to name as fast as possible 36 colours (i.e. yellow, red, green and blue) or figures (i.e. circle, square, triangle and star) in a given order presented on a sheet by six rows of six items in each. For both tasks, a composite score was calculated: the number of total correct responses was divided by total naming

time, meaning that the higher the value, the more items the child could name correctly per second, thus displaying better RAN. *Language skills* were measured using Ordforståelse [Word comprehension], a subtest from the standardised WISC-V (Wechsler, 2017), where the child needs to name pictures and define words. There are 29 items in total, and the task is stopped after three consecutive errors. For picture items (1–4), either 0 or 1 point is given, and for the rest of the items (5–29) 0, 1 or 2 points are given, based on the correctness of the definition the child gives. Therefore, the total maximum score is 54 points.

To control for demographic factors, we included children's *age* (date of birth), *gender*, *status of second language learner* (i.e. "Norwegian is a child's second language" – yes/no) and *mother's and father's highest educational level* (1=comprehensive school, 2=upper secondary school, 3=bachelor, 4=master, 5=PhD), as reported by the parents in a questionnaire.

Procedure

When needed, the test instructions were translated first into English and then into Norwegian, and the item level questions were also back-translated to ensure the correspondence and quality of the translations, either by the research team members fluent in both languages or using authorized translators. Children's mathematical and other cognitive skills were measured as part of the larger data collection in the project. Apart from Raven, which was administered in the classrooms, all tests were conducted when the children participated in one 4-h session at the data collection site, during which they were tested individually (WISC-V Arithmetic, LukiMat, WISC-V Digit and Picture Spans, RAN, Word comprehension) and in small groups (SYMP, Regnefaktaprøven) by trained research assistants. The research assistants were bachelor and master students studying (special) education. Small breaks and a lunch break were given during the testing session.

The children in Norway start schooling in August the year they turn the age of six, and teaching follows the guidelines of the national core curriculum (The Norwegian Directorate for Education & Training, 2013). Accordingly, children receive around 3 h of mathematics instruction per week, and the main foci in mathematics learning are on Numbers on number range up to 20 (e.g. counting, addition and subtraction, comparing numbers), Geometry (e.g. recognise and describe characteristics of simple two- and three-dimensional figures), Measurement (e.g. length and money) and Statistics (e.g. illustrate data using tally marks and bar graphs). The data collection took place between mid-March and the beginning of May 2019, meaning that the children had received 7–9 months of formal mathematics instruction.

The tests were scored and coded by three trained research assistants, and around 14% of the data (i.e. the data of three randomly chosen participants from each classroom for each test) were double coded by the first author. The correlations of sum scores in each test ranged between $r=0.95$ and 1.00 after double coding. When needed, the original data (paper sheets) were checked regarding the non-matching sum scores, and the final data matrices were corrected at item level accordingly.

Data analysis

The grouping on the two measures of SNMP and three measures of WM was conducted using latent class cluster analysis (LCCA) as implemented in Latent GOLD 5.1 software

(Vermunt & Magidson, 2016). The measures were first discretized to balance the distributions, and then used as clustering variables. Information criteria (e.g. AIC, AIC3, BIC, CAIC and SABIC, as implemented in Latent GOLD) were used for evaluating the appropriateness of the models, and a conditional bootstrap procedure was performed to compare sets of competing models. Based on the final model, each participant was classified into one group according to the classification probabilities, and the resulting variable was then used as a grouping variable for further analyses. Distribution of gender in different groups was examined using configural frequency analysis (CFA; Eye et al., 1996). Group differences in mathematics outcome measures were examined by a series of ANOVAs, which were then extended to ANCOVAs with measures of nonverbal intelligence, rapid automatized naming and word comprehension as well as children's age, gender, status of second language learner and mother's and father's educational level as covariates. Both sets of analyses were conducted with statistics software Jamovi v. 1.2.5.0 (The jamovi project, 2020).

Results

Descriptive statistics and correlations between all measures are given in the Appendix, Table A1 and A2, respectively. All measures showed acceptable reliability values in terms of McDonald's omega ($\omega=0.73\text{--}0.94$). Small statistically significant correlations between SNMP and WM measures were found only between 1-digit SYMP and digit backwards ($r=0.26$), 1-digit SYMP and picture span ($r=0.21$) and 2-digit SYMP and digit backwards ($r=0.15$), thus reflecting the relative independence of the two constructs. Measures of mathematics performance correlated strongly with each other ($r=0.53\text{--}0.79$), relatively weakly with digit forwards ($r=0.15\text{--}0.29$) and picture span ($r=0.15\text{--}0.26$) and moderately with digit backwards ($r=-0.38\text{--}0.45$). Both SNMP measures were also moderately associated with the measures of mathematics performance ($r=0.41\text{--}0.50$ for 1-digit SYMP and $r=0.34\text{--}0.54$ for 2-digit SYMP).

Grouping

An examination of the different information criteria from a series of LCCAs showed the lowest values (i.e. indicating best fit) for solutions with two to six classes, but them to level off between two to four classes across all indices (Table 2). This is rather expected given the different characteristics of the information criteria (Morgan, 2015). After narrowing down the most likely solutions, we used conditional bootstrapping to compare consecutive groups starting from the model with two groups. These analyses showed that the three-group solution was superior compared to the two-group solution (-2LL Diff = 30.58, $p < 0.001$), and the four-group solution superior to the three-group solution (-2LL Diff = 21.76, $p < 0.01$), but the five-group solution did not seem to significantly add to the description of the data (-2LL Diff = 14.278, $p = 0.15$). Therefore, we chose the model with four groups for further analyses.

The clustering solution explained variance significantly in all criterion variables, ranging from 25% (Backwards Digit Span) to 55% (SYMP 2-digit). In relative terms, groups one (33.6% of participants) and three (23.4%) had the lowest scores in SNMP, but compared to group one, members of group three scored significantly lower on

Table 2 Fit indices for different LCCA solutions

		BIC(LL)	AIC(LL)	AIC3(LL)	CAIC(LL)	SABIC(LL)
Model 1	1-Cluster	4152.96	4082.06	4102.06	4172.96	4089.55
Model 2	2-Cluster	4103.22	4011.05	4037.05	4129.22	4020.80
Model 3	3-Cluster	4105.91	3992.47	4024.47	4137.91	4004.47
Model 4	4-Cluster	4117.43	3982.71	4020.71	4155.43	3996.96
Model 5	5-Cluster	4136.43	3980.45	4024.45	4180.43	3996.94
Model 6	6-Cluster	4152.69	3975.43	4025.43	4202.69	3994.18
Model 7	7-Cluster	4180.27	3981.74	4037.74	4236.27	4002.73
Model 8	8-Cluster	4193.95	3974.14	4036.14	4255.95	3997.39

WM tests, particularly on Forwards Digit Span and Backwards Digit Span. Group two (25.8%) had the highest scores on both tests of SNMP, but scores close to sample mean in WM. Group four (17.2%), in contrast, performed relatively well on all measures, especially on WM and most distinctively on visual WM. Based on absolute and standardised means, the groups from one to four were labelled according to the most pronounced characteristics as (1) weak SNMP, (2) strong SNMP, (3) weak SNMP and WM, and (4) strong WM. Group differences on measures of SNMP and WM are reported in Table 3, and profiles are illustrated in Fig. 1.

Configural frequency analysis (with Lehmachers test with continuity correction) suggested some gender differences in the distribution of girls and boys into the different groups. The overall variation was significant, $\chi^2(3) = 28.72$, $p < 0.001$, and the analyses flagged for two types (i.e. frequency higher than expected by chance alone) and antitypes (i.e. frequency lower than expected by chance alone). As can be inferred from Table 4, girls were overrepresented in the weak SNMP group (type) and underrepresented in the strong SNMP group (antitype), while boys, conversely, were overrepresented in the strong SNMP group (type) and underrepresented in the weak SNMP group (antitype).

Table 3 Group differences on measures of symbolic magnitude processing and working memory

	Weak SNMP (<i>n</i> = 86)	Strong SNMP (<i>n</i> = 66)	Weak SNMP & WM (<i>n</i> = 60)	Strong WM (<i>n</i> = 44)	<i>F</i>	<i>p</i>	η^2p
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)			
SYMP 1-digit	15.27 ^a (3.30)	20.73 ^b (3.14)	14.90 ^a (3.56)	19.48 ^b (3.03)	52.33	< .001	.38
SYMP 2-digit	5.43 ^a (2.11)	12.53 (3.14)	6.20 ^a (2.47)	8.57 (2.98)	101.48	< .001	.55
Forwards digit span	7.78 ^a (0.96)	6.91 (1.49)	5.45 (1.32)	7.84 ^a (0.89)	53.33	< .001	.39
Backwards digit span	6.37 ^{ab} (1.65)	6.41 ^{ac} (1.66)	4.47 (1.63)	7.09 ^{bc} (1.43)	27.65	< .001	.25
Picture Span	15.99 ^{ab} (5.16)	16.18 ^{ac} (4.50)	13.90 ^{bc} (5.32)	27.98 (2.87)	88.65	< .001	.51

Note. Means within a row sharing the same subscripts do not differ from each other at $p < .05$ (SYMP 1-digit and backwards digit span with Bonferroni adjustment, otherwise with Games-Howell adjustment)

Fig. 1 Group profiles on symbolic numerical magnitude processing and working memory

Table 4 Configural frequency analysis of gender distribution in different groups

Configuration (group gender)	<i>fo</i>	<i>fe</i>	χ^2	<i>p</i>	
11	53	39.641	3.407	.00033	Type
12	33	46.359	−3.407	.00033	Antitype
21	17	30.422	−3.697	.00011	Antitype
22	49	35.578	3.697	.00011	Type
31	20	27.656	−2.114	.01726	
32	40	32.344	2.114	.01726	
41	28	20.281	2.394	.00832	
42	16	23.719	−2.394	.00832	

Note. *fo* frequency observed, *fe* frequency expected; gender coded: 1 = girl, 2 = boy

Mathematics performance by groups

Next, group differences on mathematics outcome measures were examined. Again, the grouping explained individual differences in the target variables significantly, with explained variance ranging from 21% (word problem-solving) to 30% (counting). Groups strong SNMP and strong WM did not differ from each other in any of the measures, and the weak SNMP group and the weak SNMP and WM group did not differ from each other in addition and subtraction facts. Other than that, pairwise group differences were all significant (see Table 5). The strong SNMP and strong WM groups

received the highest scores on all measures, followed by both the weak SNMP group and the weak SNMP and WM group on addition and subtraction facts, and first by the weak SNMP and then by the weak SNMP and WM group on both counting and word problem-solving.

In order to control for the effects of factors that likely contribute to children's mathematics performance (i.e. domain-general skills, language and some key demographics), we ran a series of ANCOVAs. Group differences remained significant (Table 6), yet the explained variance somewhat reduced, ranging now from 7% (word problem-solving) to 16% (addition facts). Gender ($F=6.39$, $p=0.012$, $\eta^2p=0.03$), fluid intelligence ($F=5.27$, $p=0.023$, $\eta^2p=0.03$) and RAN Colours ($F=6.32$, $p=0.013$, $\eta^2p=0.03$) had a significant effect on counting; word comprehension ($F=25.30$, $p<0.001$, $\eta^2p=0.12$) on word problem-solving; and RAN Figures on both addition ($F=4.76$, $p=0.030$, $\eta^2p=0.02$) and subtraction facts ($F=5.77$, $p=0.017$, $\eta^2p=0.03$) (see Table 7).

An inspection of adjusted means showed the covariates to moderate group differences so that groups weak SNMP and strong WM did not differ from each other anymore in counting skills. In word problem-solving, only the strong SNMP and the weak SNMP and WM groups differed from each other, and in both addition and subtraction facts, only the

Table 5 Group differences on mathematics outcome measures

	Weak SNMP ¹ (<i>n</i> =86)	Strong SNMP (<i>n</i> =66)	Weak SNMP & WM ² (<i>n</i> =60)	Strong WM ³ (<i>n</i> =44)	<i>F</i>	<i>p</i>	η^2p
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)			
Counting	18.67 (6.01)	23.39 ^a (3.97)	15.18 (5.78)	23.18 ^a (3.22)	35.35	<.001	.30
Word problem-solving	10.79 (2.73)	12.74 ^a (3.17)	9.13 (2.87)	13.09 ^a (3.35)	21.70	<.001	.21
Addition facts	4.35 ^a (2.80)	8.59 ^b (4.82)	3.20 ^a (2.73)	7.86 ^b (4.76)	30.03	<.001	.26
Subtraction facts	3.27 ^a (2.69)	6.95 ^b (4.11)	2.44 ^a (2.33)	6.26 ^b (4.25)	27.26	<.001	.25

Note. Means within a row sharing the same subscript do not differ from each other at $p<.05$ (with Bonferroni adjustment). ¹In Word problems and Addition facts $n=85$; ²in Subtraction facts $n=59$; ³in Subtraction facts $n=42$

Table 6 Adjusted group differences on mathematics outcome measures (after controlling for age, gender, parents' educational level, fluid intelligence, RAN, and word comprehension)

	Weak SNMP (<i>n</i> =86)	Strong SNMP (<i>n</i> =66)	Weak SNMP & WM (<i>n</i> =60)	Strong WM (<i>n</i> =44)	<i>F</i>	<i>p</i>	η^2p
	<i>M</i> (<i>se.</i>)	<i>M</i> (<i>se.</i>)	<i>M</i> (<i>se.</i>)	<i>M</i> (<i>se.</i>)			
Counting	19.40 ^{ab} (0.60)	22.61 ^{ac} (0.72)	16.46 ^{bc} _d (0.78)	21.33 _d (0.87)	10.37	<.001	.14
Word problem-solving	11.16 (0.35)	12.47 ^a (0.42)	10.07 ^a (0.46)	11.98 (0.51)	4.68	.004	.07
Addition facts	4.545 ^a (0.46)	8.53 ^{ab} (0.55)	3.94 ^b (0.60)	6.41 (0.67)	11.96	<.001	.16
Subtraction facts	3.51 ^a (0.40)	6.71 ^{ab} (0.49)	3.35 ^b (0.53)	5.22 (0.60)	9.09	<.001	.13

Note. Means within a row sharing the same subscript are significantly different at the $p<.05$ level (with Bonferroni adjustment). *N* in Counting and Word problem-solving: total=201, Weak SNMP=68, Strong SNMP=50, Weak SNMP & WM=46, Strong WM=37; in Addition facts: total=200, Weak SNMP=67, Strong SNMP=50, Weak SNMP & WM=46, Strong WM=37; in Subtraction facts: total=198, Weak SNMP=68, Strong SNMP=50, Weak SNMP & WM=45, Strong WM=35

Table 7 Effects of covariates on target measures

	Counting (<i>n</i> = 201)			Word problem–solv- ing (<i>n</i> = 201)			Addition facts (<i>n</i> = 200)			Subtraction facts (<i>n</i> = 198)		
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	η^2p
Age	0.06	.799	.00	0.57	.449	.00	0.34	.562	.00	1.83	.178	.01
Gender	6.39	.012	.03	2.62	.107	.01	3.41	.067	.02	3.20	.075	.02
Mother's educ.level	3.15	.077	.02	0.14	.711	.00	0.92	.338	.00	2.09	.149	.01
Father's educ. level	2.56	.111	.01	0.03	.854	.00	2.11	.148	.01	0.82	.367	.00
Second lang. status	3.73	.055	.02	3.12	.079	.02	0.83	.363	.00	1.09	.298	.01
Fluid intelligence	5.27	.023	.03	2.95	.087	.02	0.52	.470	.00	0.08	.784	.00
RAN Colours	6.32	.013	.03	0.76	.384	.00	3.43	.066	.02	3.03	.083	.02
RAN Figures	0.30	.582	.00	1.97	.162	.01	4.76	.030	.02	5.77	.017	.03
Word comprehension	3.11	.080	.02	25.30	<.001	.12	1.97	.162	.01	2.07	.152	.01

weak SNMP and strong SNMP groups, and the strong SNMP and the weak SNMP and WM groups, differed from each other, respectively. These results, however, need to be interpreted with caution, due to considerable missing data in some of the covariates.¹

Discussion

The aim of this study was to investigate the patterning of symbolic numerical magnitude processing (SNMP) and working memory (WM) by identifying different performance profiles among first graders, and, furthermore, how those profiles predict children's performance in mathematics. The novelty of this study lays in (i) examining the two constructs, SNMP and WM, simultaneously, (ii) including three components of WM (verbal, visual and central executive), (iii) extracting empirical performance profiles of SNMP and WM by means of latent class clustering and (iv) predicting mathematics performance while controlling for differences in key demographics and cognitive factors. The main findings, which are elaborated in the following, were as follows: (i) four different profiles of SNMP and WM performance were identified among first graders with children displaying relative strengths in SNMP and WM, and (ii) SNMP and WM were shown to have a unique, even compensatory role in children's mathematics performance, yet their individual contribution might depend on the types of mathematical skill in question (e.g. arithmetic facts vs. word problem–solving).

Four groups of children with different profiles in their performance on SNMP and WM were identified. Congruent with previous studies (Kroesbergen & van Dijk, 2015; Toll et al., 2016), we found a group with weak SNMP skills (33.6%) and a group with weak SNMP and WM (23.4%). We did not find a group with weak WM skills only, which is

¹ This was especially evident with key demographics, as some parents did not return the questionnaire with complete information, thus resulting in an effective numerus of 217 for these analyses. Additional non-presence occurred in individual sessions leading to a share of missing values ranging from 16 to 24%. Note, however, that the attrition rate was rather equal across the different groups (Group 1: 21%; Group 2: 24%; Group 3: 23%; and Group 4: 16%).

different from previous studies using a priori defined cut-off scores (Kroesbergen & van Dijk, 2015; Toll et al., 2016). Note, however, that even then, Toll et al. (2016) reported only relatively few children (5-year-olds) in the weak WM group compared to other groups. Furthermore, in contrast to previous studies (Kroesbergen & van Dijk, 2015; Toll et al., 2016), which treated children having no weakness in either NMP or WM as one group, our clustering revealed two separate groups, one characterized by strong SNMP (25.8%) and another group by strong WM, especially visual WM (17.2%). Moreover, girls were slightly overrepresented in the weak SNMP group and boys in the strong SNMP group, while in Toll et al.'s study (2016), no gender differences were detected. The type of tasks used for measuring SNMP in each study might partly explain this difference.

When comparing our results to previous research, two differing aspects need to be noted. Firstly, as mentioned before, prior studies (Kroesbergen & van Dijk, 2015; Toll et al., 2016) used a priori defined cut-off points to classify children into different groups and treated all those with performance above the 25th percentile on both NMP and WM as one group (i.e. without weakness). The empirically derived classification in the present study may thus provide a different, less artificial, in a sense, account of the patterning of SNMP and WM among children, which, consequently, also influences the predictions on mathematics performance. Secondly, previous studies used only visuospatial WM in the grouping as an indicator of WM, whereas we included three WM components in our study. As the correlations between the WM components were not particularly strong ($r < 0.30$), thus reflecting some degree of independence, those were treated as separate indicators in the clustering. The patterning of the different components of WM within each group was nevertheless rather similar, with the exceptions of the weak SNMP and WM group displaying relatively low verbal WM and central executive in relation to other groups, and the strong WM skills group exhibiting the most distinctive level of visual WM. To sum, the use of latent class clustering enabled us to identify four groups with different performance profiles on SNMP and WM among first graders; two groups similar to previous research—that is the weak SNMP group and the weak SNMP and WM group—and two additional groups characterized by either strong SNMP or strong WM.

When group differences in mathematics outcome measures (i.e. verbal counting skills, word problem-solving, addition and subtraction facts) were examined, the grouping explained variance ranging from 21 to 30%. In a sense, strong SNMP or WM skills seem to be compensatory in relation to mathematics performance, as those groups (i.e. strong SNMP and strong WM) outperformed the others, but did not differ from each other on mathematics performance. Previous research has also highlighted that good SNMP and WM skills are often conjoined with good mathematics performance. For example, SNMP has been connected with arithmetic fluency (Vanbinst et al., 2012) and WM with word problem-solving (Fung & Swanson, 2017). As in previous research (Kroesbergen & van Dijk, 2015; Toll et al., 2016), in our study, weakness in either SNMP or both in SNMP and WM was related to lower mathematics performance. The weak SNMP group and the weak SNMP and WM group were inferior in addition and subtraction facts compared to groups with better SNMP skills. However, there was no difference between these two groups, meaning that already weak SNMP skills alone were a sufficient condition for lower performance in arithmetic facts. The results regarding counting and word problems were similar to those on arithmetic facts although, here, the weak SNMP group outperformed the weak SNMP and WM group, suggesting that in tasks of counting and word problem-solving, WM skills may compensate for weak SNMP skills. This is in line with previous research showing a connection between weak WM capacity and inferior performance in counting (Preßler et al., 2013).

The inclusion of covariates reduced the explained variance to a range of 7 to 16%. Even though the covariates individually had a relatively small impact on mathematics outcome measures, they did moderate group differences especially on arithmetic facts and word problem-solving. The strongest effect ($\eta^2p=0.12$) was found from word comprehension on word problems, which was expected, as language skills (both receptive and expressive) are required in word problem-solving tasks. That is, the child needs to both understand the meaning of the (mathematical) words given verbally to them and express the answer for the problem orally. Hence, in word problem-solving, the only group difference was now found between the strong SNMP group and the weak SNMP and WM group, favouring the former. In contrast to previous research (Kroesbergen & van Dijk, 2015), where differences in word comprehension skills were not taken into account, the group having weakness only in SNMP did not perform worse in word problem-solving than the groups without such weakness. In arithmetic facts, the results were similar to those without covariates, except that no differences were found between the strong WM group and the groups having at least one weakness. This further supports the inference that individual differences in SNMP contribute to arithmetic facts. In counting, the group differences remained the same, except that the weak SNMP and strong WM groups did not differ anymore in their performance.

In addition to contributing to research on the roles of a domain-specific (i.e. SNMP) and a domain-general (i.e. WM) factors as precursors of early mathematics performance, the present findings also have implications for the discussion on the extent to which underlying factors could or should be emphasized in early screening (e.g. for identifying children at risk for mathematical learning difficulties). Based on our results, assessing children's SNMP skills already in the first grade might be sufficient for such screening. Using a valid and reliable measure of SNMP, such as SYMP (Brankaer et al., 2017), which can be administered effortlessly by classroom teachers repeatedly over the school year, might serve as a good screening tool for this purpose. This suggestion is supported also by previous studies showing children with mathematical difficulties to perform lower in SNMP tasks compared to their peers without such difficulties (e.g. Brankaer et al., 2017; Landerl et al., 2009).

Limitations and future directions

Due to being cross-sectional, the findings of our study on the patterning of SNMP and WM and their predictions on early mathematics performance can only be considered correlational, not causal. However, we believe they set a good foundation for further hypotheses that could be tested within a longitudinal design, both in terms of the stability and change in the given patterning (i.e. qualitative changes in profiles or shifts in group membership) and the developmental trajectories or long-term predictions over time. Our measurements were somewhat limited in the sense that single measures were used to represent different constructs, which was mostly due to practical reasons. The children already completed numerous tasks and tests, including the ones we used as covariates in the study. Future studies would nevertheless benefit from including more than one measure for different constructs to enhance validity and to better take into account unreliability and sources of variation. Another important limitation was the missing information on key demographics, due to which the number of cases in the analyses of covariance was reduced. Because of this, the effects of covariates and how they moderated differences in mathematics outcome measures must be interpreted with caution. Also, the fact that our sample of children was somewhat selective in the sense

that they came mainly from families of middle or high educational level must be taken into consideration, even though it is not clear whether this, in fact, would have any influence on the relations, and hence also predictions, between the target constructs. To better account for the role of familial background, a more heterogeneous sample would be needed, and perhaps also a better proxy for SES (e.g. yearly income of household).

Conclusions

Our empirical approach enabled us to identify four groups of children with different patterns of SNMP and WM performance. Two groups resembled those found in previous studies (the weak SNMP group and the weak SNMP and WM group), while two were somewhat unique (the strong SNMP group and the strong WM group), which potentially provides us with new insights into the complexity of factors underlying the development of children's mathematical skills. Our findings suggest that WM and particularly SNMP may play an important role in mathematics performance, even after taking into account various key demographics and cognitive factors, and that their contribution might be partly compensatory. As an implication for practice, screening SNMP skills already in the first grade might be an effective and feasible approach to identify children with potential problems in mathematics learning and at risk for mathematical learning difficulties later on.

Appendix A

Table A1 Descriptive Statistics of Variables

	<i>n</i>	<i>M</i>	SD	min.	max.	skewness	kurtosis	ω
1. SYMP 1-digit	256	17.31	4.15	7	29	0.03	-0.18	.88
2. SYMP 2-digit	256	7.98	3.91	1	21	0.75	0.11	.86
3. WM: Digit forwards	256	7.02	1.52	0	11	-0.89	3.37	.76
4. WM: Digit backwards	256	6.06	1.85	0	11	-0.23	0.86	.73
5. WM: Picture span	256	17.61	6.72	2	34	0.31	-0.54	.83
6. Counting	256	19.85	6.03	4	29	-0.65	-0.43	.90
7. Word problem solving	255	11.30	3.34	2	22	-0.03	0.28	.79
8. Addition facts	255	5.78	4.38	0	26	1.37	2.81	.93
9. Subtraction facts	253	4.53	3.82	0	22	1.32	2.45	.89
10. Fluid intelligence	239	23.78	5.42	8	34	-0.32	-0.41	.85
11. Word comprehension	255	13.16	3.22	3	24	-0.06	0.98	.73
12. RAN colors	255	0.98	0.25	0.21	1.89	0.08	0.80	.94
13. RAN figures	255	0.73	0.22	0.23	1.80	0.65	2.13	.94

Note. ω = McDonald's omega for reliability, *** $p < .001$, ** $p < .01$, * $p < .05$

Table A2 Correlations between Variables

	1	2	3	4	5	6	7	8	9	10	11	12
1. SYMP 1-digit	–											
2. SYMP 2-digit	0.55***	–										
3. WM: Digit forwards	0.09	0.01	–									
4. WM: Digit backwards	0.26***	0.15*	0.29***	–								
5. WM: Picture span	0.21***	0.08	0.27***	0.23***	–							
6. Counting	0.47***	0.47***	0.29***	0.45***	0.26***	–						
7. Word problem solving	0.41***	0.34***	0.29***	0.38***	0.32***	0.55***	–					
8. Addition facts	0.49***	0.54***	0.18***	0.43***	0.21***	0.61***	0.53***	–				
9. Subtraction facts	0.50***	0.49***	0.15*	0.39***	0.15*	0.60***	0.53***	0.79***	–			
10. Fluid intelligence	0.29***	0.22***	0.22***	0.36***	0.35***	0.39***	0.34***	0.29***	0.21***	–		
11. Word comprehension	0.16*	–0.02	0.21***	0.23***	0.28***	0.26***	0.45***	0.23***	0.25***	0.24***	–	
12. RAN colors	0.46***	0.23***	0.20**	0.35***	0.21***	0.36***	0.27***	0.38***	0.35***	0.28***	0.16*	–
13. RAN figures	0.36***	0.20**	0.19**	0.35***	0.24***	0.30***	0.30***	0.37***	0.34***	0.27***	0.25***	0.54***

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

Acknowledgements We thank all the participating children, their teachers and parents, as well as the research assistants involved in the data collection.

Author contribution Conceptualization: Riikka Mononen, Markku Niemivirta; methodology: Riikka Mononen, Markku Niemivirta; formal analysis: Markku Niemivirta; investigation: Riikka Mononen; writing—original draft preparation: Riikka Mononen, Markku Niemivirta; funding acquisition: Riikka Mononen; resources: Riikka Mononen.

Funding Open access funding provided by University of Oslo (incl Oslo University Hospital). This work was supported by the Norwegian Research Council [Grant number: 283396], for the first author.

Availability of data and materials Upon request from the correspondent author.

Code availability Not applicable.

Declarations

Ethics approval Norwegian Centre for Research Data.

Consent for participation Written consent from parents and teachers.

Consent for publication Written consent from parents and teachers.

Conflict of interest Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, 106(1), 20–29. <https://doi.org/10.1016/j.jecp.2009.11.003>
- Andersson, U. (2008). Working memory as a predictor of written arithmetical skills in children: The importance of central executive functions. *British Journal of Educational Psychology*, 78(2), 181–203. <https://doi.org/10.1348/000709907X209854>
- 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>
- Baddeley, A. (2010). Working memory. *Current Biology*, 20(4), R136–R140. <https://doi.org/10.1016/j.cub.2009.12.014>
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- Brankaer, C., Ghesquière, P., & De Smedt, B. (2017). Symbolic magnitude processing in elementary school children: A group administered paper-and-pencil measure (SYMP Test). *Behavior Research Methods*, 49(4), 1361–1373. <https://doi.org/10.3758/s13428-016-0792-3>
- Butterworth, B. (2005). Developmental dyscalculia. In J. I. D. Campbell (Ed.), *Handbook of Mathematical Cognition* (pp. 455–467). Psychology Press.

- Cañizares, D. C., Crespo, V. R., & Alemañy, E. G. (2012). Symbolic and non-symbolic number magnitude processing in children with developmental dyscalculia. *The Spanish Journal of Psychology*, 15(3), 952–966. <https://doi.org/10.1016/j.jecp.2008.12.006>
- Chan, W. W. L., & Wong, T. T.-Y. (2019). Visuospatial pathways to mathematical achievement. *Learning and Instruction*, 62, 11–19. <https://doi.org/10.1016/j.learninstruc.2019.03.001>
- David, C. V. (2012). Working memory deficits in math learning difficulties: A meta-analysis. *International Journal of Developmental Disabilities*, 58(2), 67–84. <https://doi.org/10.1179/2047387711Y.0000000007>
- 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>
- 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>
- DeStefano, D., & LeFevre, J.-A. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology*, 16(3), 353–386. <https://doi.org/10.1080/09541440244000328>
- Eye, A., Spiel, C., & Wood, P. K. (1996). Configural frequency analysis in applied psychological research. *Applied Psychology: An International Review*, 45(4), 301–327. <https://doi.org/10.1111/j.1464-0597.1996.tb00770.x>
- Feigenson, L., Libertus, M. E., & Halberda, J. (2013). Links between the intuitive sense of number and formal mathematics ability. *Child Development Perspectives*, 7(2), 74–79. <https://doi.org/10.1111/cdep.12019>
- Friso-van den Bos, I., van der Ven, S. H. G., Kroesbergen, E. H., & van Luit, J. E. H. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review*, 10, 29–44. <https://doi.org/10.1016/j.edurev.2013.05.003>
- Fung, W., & Swanson, H. L. (2017). Working memory components that predict word problem solving: Is it merely a function of reading, calculation, and fluid intelligence? *Memory & Cognition*, 45(5), 804–823. <https://doi.org/10.3758/s13421-017-0697-0>
- Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities*, 37(1), 4–15. <https://doi.org/10.1177/00222194040370010201>
- Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, M. C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology*, 88(2), 121–151. <https://doi.org/10.1016/j.jecp.2004.03.002>
- Goffin, C., & Ansari, D. (2019). How are symbols and nonsymbolic numerical magnitudes related? Exploring bidirectional relationships in early numeracy. *Mind, Brain, and Education*, 13(3), 143–156. <https://doi.org/10.1111/mbe.12206>
- Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology*, 103(1), 17–29. <https://doi.org/10.1016/j.jecp.2008.04.001>
- Karagiannakis, G., Baccaglini-Frank, A., & Papadatos, Y. (2014). Mathematical learning difficulties subtypes classification. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00057>
- Klausen, T., & Reikerås, E. (2016). *Regnefaktaprøven*. Lesesenteret, University of Stavanger, Norway.
- Koponen, T., Eklund, K., & Salmi, P. (2018). Cognitive predictors of counting skills. *Journal of Numerical Cognition*, 4(2), 410–428. <https://doi.org/10.5964/jnc.v4i2.116>
- Kroesbergen, E. H., & van Dijk, M. (2015). Working memory and number sense as predictors of mathematical (dis-)ability. *Zeitschrift Für Psychologie*, 223(2), 102–109. <https://doi.org/10.1027/2151-2604/a000208>
- Kyttälä, M., Kanerva, K., Munter, I., & Björn, P. M. (2019). Working memory resources in children: Stability and relation to subsequent academic skills. *Educational Psychology*, 39(6), 709–728. <https://doi.org/10.1080/01443410.2018.1562046>
- Lafay, A., St-Pierre, M.-C., & Macoir, J. (2019). Impairment of non-symbolic number processing in children with mathematical learning disability. *Journal of Numerical Cognition*, 5(1), 86–104. <https://doi.org/10.5964/jnc.v5i1.177>
- Landerl, K., Fussnegg, 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>

- Li, Y., Zhang, M., Chen, Y., Deng, Z., Zhu, X., & Yan, S. (2018). Children's non-symbolic and symbolic numerical representations and their associations with mathematical ability. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.01035>
- Menon, V. (2016). Working memory in children's math learning and its disruption in dyscalculia. *Current Opinion in Behavioral Sciences*, 10, 125–132. <https://doi.org/10.1016/j.cobeha.2016.05.014>
- Morgan, G. B. (2015). Mixed mode latent class analysis: An examination of fit index performance for classification. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(1), 76–86. <https://doi.org/10.1080/10705511.2014.935751>
- Nuerk, H. C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition*, 82(1), B25–33. [https://doi.org/10.1016/S0010-0277\(01\)00142-1](https://doi.org/10.1016/S0010-0277(01)00142-1)
- Passolunghi, M. C., & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology*, 88(4), 348–367. <https://doi.org/10.1016/j.jecp.2004.04.002>
- Passolunghi, M. C., Cargnelutti, E., & Pastore, M. (2014). The contribution of general cognitive abilities and approximate number system to early mathematics. *British Journal of Educational Psychology*, 84(4), 631–649. <https://doi.org/10.1111/bjep.12054>
- Preßler, A.-L., Krajewski, K., & Hasselhorn, M. (2013). Working memory capacity in preschool children contributes to the acquisition of school relevant precursor skills. *Learning and Individual Differences*, 23, 138–144. <https://doi.org/10.1016/j.lindif.2012.10.005>
- Purpura, D. J., & Ganley, C. M. (2014). Working memory and language: Skill-specific or domain-general relations to mathematics? *Journal of Experimental Child Psychology*, 122, 104–121. <https://doi.org/10.1016/j.jecp.2013.12.009>
- Raven, J. C., Court, J. H., & Raven, J. (1990). *Manual for Raven's progressive matrices and vocabulary scales: Section 2, Coloured progressive matrices*. Oxford University Press.
- Reynvoet, B., & Brysbaert, M. (1999). Single-digit and two-digit Arabic numerals address the same semantic number line. *Cognition*, 72(2), 191–201. [https://doi.org/10.1016/S0010-0277\(99\)00048-7](https://doi.org/10.1016/S0010-0277(99)00048-7)
- 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(3), 361–395. <https://doi.org/10.1016/j.cognition.2006.01.005>
- Salminen, J., & Koponen, T. (2011). *Oppimisen arviointi: Matematiikan oppimisen seurannan välineet. Käsitkirja [LukiMat – Assessment of learning: Progress monitoring of mathematics learning. Manual]*. Niilo Mäki Institute.
- Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., & Smedt, B. D. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science*, 20(3), e12372. <https://doi.org/10.1111/desc.12372>
- Semel, E., Wiig, E., & Secord, W. (2003). *Clinical Evaluation of Language Fundamentals – Fourth Edition (CELF-4)*. Norwegian version. Pearson.
- Siemann, J., & Petermann, F. (2018). Innate or acquired? - Disentangling number sense and early number competencies. *Frontiers in Psychology*, 9, 571. <https://doi.org/10.3389/fpsyg.2018.00571>
- The jamovi project. (2020). *Jamovi (Version 1.2) [Computer Software]*. <https://www.jamovi.org>
- The Norwegian Directorate for Education and Training. (2013). *Curriculum for the Common Core Subject of Mathematics (MAT1–04)*. <https://www.udir.no/laring-og-trivsel/lareplanverket/finn-lareplan/lareplan/>. Accessed 20 June 2020.
- Toll, S. W. M., Kroesbergen, E. H., & Luit, J. E. H. V. (2016). Visual working memory and number sense: Testing the double deficit hypothesis in mathematics. *British Journal of Educational Psychology*, 86(3), 429–445. <https://doi.org/10.1111/bjep.12116>
- Vanbinst, K., Ghesquière, P., & Smedt, B. D. (2012). Numerical magnitude representations and individual differences in children's arithmetic strategy use. *Mind, Brain, and Education*, 6(3), 129–136. <https://doi.org/10.1111/j.1751-228X.2012.01148.x>
- Vermunt, J. K., & Magidson, J. (2016). *Upgrade Manual for Latent GOLD 5.1*. Belmont Massachusetts: Statistical Innovations Inc.
- Wechsler, D. (2017). *WISC-V Wechsler Intelligence Scale for Children – fifth edition. Norwegian version*. Pearson.
- Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, 91(3), 415–438. <https://doi.org/10.1037/0022-0663.91.3.415>
- Zheng, X., Swanson, H. L., & Marcoulides, G. A. (2011). Working memory components as predictors of children's mathematical word problem solving. *Journal of Experimental Child Psychology*, 110(4), 481–498. <https://doi.org/10.1016/j.jecp.2011.06.001>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Current themes of research

Riikka Mononen: Individual differences in children's mathematics development, its interplay with other cognitive skills, motivation, and emotions; mathematical learning difficulties; assessment and intervention in mathematics.

Markku Niemivirta: Developmental relationships between motivation, learning, and well-being; situational dynamics between motivation and performance; individual and contextual predictors of motivation, particularly in mathematics and from a developmental perspective.

Most relevant publications in the field of Psychology of Education

Riikka Mononen:

Lopez-Pedersen, A., Mononen, R., Korhonen, J., Aunio, P., & Melby-Lervåg, M. (2020). Validation of an early numeracy screener for first graders. *Scandinavian Journal of Educational Research*, 65(3).

Aunio, P., Korhonen, J., Ragpot, L., Törmänen, M., Mononen, R., & Henning, E. (2019). Multi-factorial approach to early numeracy – the effects of cognitive skills, language factors and kindergarten attendance on early numeracy performance of South African first graders. *International Journal of Educational Research*, 97, 65–76.

Funderud, T., Mononen, R., Radišić, J., & Laine, A. (2018). A comparative study of variations in arithmetic fluency between Norwegian and Finnish third graders. *European Journal of Special Needs Education*, 34(5), 572–585.

Markku Niemivirta:

Nuutila K., Tapola, A., Tuominen, H., Kupiainen, S., Pásztor, A., & Niemivirta, M. (2020). Reciprocal predictions between interest, self-efficacy, and performance during a task. *Frontiers in Education*, 5:36.

Tuominen, H., Niemivirta, M., Lonka, K., & Salmela-Aro, K. (2020). Motivation across a transition: Changes in achievement goal orientations and academic well-being from elementary to secondary school. *Learning and Individual Differences*, 79.

Niemivirta, M., Pulkka, A.-T., Tapola, A., & Tuominen, H. (2019). Achievement goal orientations: A person-oriented approach. In K. A. Renninger, & S. E. Hidi (Eds.), *The Cambridge Handbook of Motivation and Learning* (pp. 566–616). Cambridge: Cambridge University Press.

Aunio, P. & Niemivirta, M. (2010). Predicting children's mathematical performance in grade one by early numeracy. *Learning and Individual Differences*, 20(5), 427–435.

Boekaerts, M. & Niemivirta, M. (2000). Self-regulated learning: Finding a balance between learning- and ego-protective goals. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 417–450). San Diego, CA: Academic Press.