

Reliability and Validity of Nonsymbolic and Symbolic Comparison Tasks in School-Aged Children

Danilka Castro¹, Nancy Estévez², David Gómez¹ and Pablo Ricardo Dartnell¹

¹ *Centro de Investigación Avanzada en Educación (Chile)*

² *Centro de Neurociencias de Cuba (Cuba)*

Abstract. Basic numerical processing has been regularly assessed using numerical nonsymbolic and symbolic comparison tasks. It has been assumed that these tasks index similar underlying processes. However, the evidence concerning the reliability and convergent validity across different versions of these tasks is inconclusive. We explored the reliability and convergent validity between two numerical comparison tasks (nonsymbolic vs. symbolic) in school-aged children. The relations between performance in both tasks and mental arithmetic were described and a developmental trajectories' analysis was also conducted. The influence of verbal and visuospatial working memory processes and age was controlled for in the analyses. Results show significant reliability ($p < .001$) between Block 1 and 2 for nonsymbolic task (global adjusted RT (adjRT): $r = .78$, global efficiency measures (EMs): $r = .74$) and, for symbolic task (adjRT: $r = .86$, EMs: $r = .86$). Also, significant convergent validity between tasks ($p < .001$) for both adjRT ($r = .71$) and EMs ($r = .70$) were found after controlling for working memory and age. Finally, it was found the relationship between nonsymbolic and symbolic efficiencies varies across the sample's age range. Overall, these findings suggest both tasks index the same underlying cognitive architecture and are appropriate to explore the Approximate Number System (ANS) characteristics. The evidence supports the central role of ANS in arithmetic efficiency and suggests there are differences across the age range assessed, concerning the extent to which efficiency in nonsymbolic and symbolic tasks reflects ANS acuity.

Received 31 October 2016; Revised 4 November 2017; Accepted 7 November 2017

Keywords: ANS, arithmetic, reliability, validity, working memory.

The Approximate Number System (ANS) is a foundational cognitive system that allows us to represent the cardinality of sets of objects in an analogue format. Due to the imprecise nature of the ANS, it has been suggested that mental representations of quantities are oriented in a “mental number line” (Dehaene & Changeux, 1993; Whalen, Gallistel, & Gelman, 1999) increasingly overlapping with increasing numerosity in a way that performance improves when the distance between the numerosities to be compared is larger, or more precisely: Follows the Weber's law: the extent to which two stimuli can be discriminated depends upon the ratio between them (Izard & Dehaene, 2007). Several studies support the idea that humans rely on number sense in order to solve certain tasks involving symbolic stimuli. This implies the existence of an interface between the system of verbal numerals and the analogue representations system (non-verbal) (Izard & Dehaene, 2007; Rousselle & Noël, 2007).

Following this idea, it has been suggested that the ANS plays an important role in the emergence of

individual differences in math performance (Desoete, Ceulemans, De Weerd, & Pieters, 2012; Halberda, Ly, Wilmer, Naiman, & Germine, 2012; Libertus, Feigenson, & Halberda, 2011; Mazzocco, Feigenson, & Halberda, 2011). For this reason, the development of this system has received special attention, both concerning the study of typical development of numerical processing (Castro, Estévez, & Pérez, 2011; Halberda, Mazzocco, & Feigenson, 2008; Landerl & Kölle, 2009) and the study of mathematical learning disabilities (Castro, Reigosa, & González, 2012; Iuculano, Tang, Hall, & Butterworth, 2008; Mussolin, Mejias, & Noël, 2010; Rousselle & Noël, 2007).

Four indices: Accuracy, Weber fraction, numerical ratio and distance effects, have been commonly used to study the ANS, on the assumption they all assess the acuity of mental representations on nonsymbolic numerosities. Accuracy is usually calculated as the proportion of trials the subjects answer correctly (Inglis & Gilmore, 2014). The Weber fraction is proposed to reflect the precision of mental representations and is estimated (on the assumption that individual accuracy

Correspondence concerning this article should be addressed to Danilka Castro Cañizares. Área de Investigación de Neurociencia y Cognición del Centro de Investigación Avanzada en Educación. Santiago (Chile).

E-mail: danilka.castro@ciae.uchile.cl

Funding from PIA-CONICYT Basal Funds for Centers of Excellence Project FB0003 is gratefully acknowledged.

How to cite this article:

Castro, D., Estévez, N., Gómez, D., & Dartnell, P. R. (2017). Reliability and validity of nonsymbolic and symbolic comparison tasks in school-aged children. *The Spanish Journal of Psychology*, 20, e75. Doi:10.1017/sjp.2017.68

on a nonsymbolic comparison task depends upon the numerosities to be compared and the precision of the corresponding numerical representations) as the value that best fit the behavioral data (Halberda & Feigenson, 2008; Inglis & Gilmore, 2014). The numerical ratio effect (NRE) (De Smedt & Gilmore, 2011; Iuculano et al., 2008; Mussolin et al., 2010; Rousselle & Noël, 2007) refers to an increase in reaction time (RT) and a decrease in accuracy in numerical discrimination when the ratio between the numerosities to be compared increasingly approaches the value of 1 (e.g., it's easier to compare 5 to 10: ratio = .5 than 8 to 10: ratio = .8). The numerical distance effect (NDE) refers to a decrease in RT during numerical discrimination with increasing numerical distance between the numerosities to be compared (e.g., it's easier to compare 2 to 6 than 2 to 3; Rousselle & Noël, 2007).

Multiple numerical comparison tasks, either in nonsymbolic (e.g., dot sets) or symbolic (e.g., Arabic digits) format, have been used to elicit the numerical processing effects described above. Different versions of these tasks have been used: intermixed presentation (dots of two different colors, intermixed but non-overlapping in the same stimulus (Halberda et al., 2008); paired presentation (pairs of sets of dots or Arabic digits are simultaneously presented; Holloway & Ansari, 2009; Maloney, Risko, Preston, Ansari, & Fugelsang, 2010; Rousselle & Noël, 2007) and sequential presentation of two stimuli (Halberda et al., 2008; Holloway & Ansari, 2009). In general, it could be assumed that performance in numerical comparison tasks (although the tasks are presented in different formats) is comparable; since it should express similar underlying processes. However, although these tasks have been used to draw conclusions on the development of numerical mental representations during childhood (either based upon the aforementioned effects, accuracy or the Weber fraction), there is debate concerning whether the different measures are reliable indices of the ANS acuity. In fact, data concerning the reliability of these measures (degree to which the repeated application of a task to the same subject produces the same results- American Educational Research Association, American Psychological Association y National Council on Measurement in Education, 1999) is rarely reported. Reliability analyses are particularly relevant because when a measurement is unreliable it reduces the likelihood of detecting differences between groups, based on this measure, even if those differences do exist.

The issue of reliability has been primarily examined through test-retest assessments, correlating data obtained in the first and second half of an experiment (e.g., Waechter, Stolz, & Besner, 2010). If a measure is reliable, the data obtained in a block of stimuli (e.g., first half of the experiment) will be highly predictive of the

results obtained in another block of stimuli (e.g., second half of the experiment).

Following this idea, Maloney et al. (2010) evaluated the test-retest reliability of the numerical distance effect in a sample of 48 adults in nonsymbolic and symbolic numerical comparison tasks. The results showed the numerical distance effect is a reliable measure in both tasks' formats; being much more reliable the numerical distance effect measurements obtained in the nonsymbolic task. Using a similar procedure, Sasanguie, Defever, van den Bussche, and Reynvoet (2011) evaluated 47 adults, but only in nonsymbolic tasks. They found significant NDE reliability levels in the paired comparison and same-different tasks, but not in the priming tasks; which did not correlate with the former tasks. Thus, the authors suggested priming tasks should be used with caution when assessing the ANS. In 2011, Gilmore, Attridge, and Inglis, evaluated 101 adults with different ANS measures, using tasks presented in nonsymbolic and symbolic format. The results showed significant reliability for all tasks administered (nonsymbolic and symbolic). Similar results were obtained by Price, Palmer, Battista, and Ansari (2012), when studying Weber fraction and the numerical ratio effect reliability elicited by three versions of nonsymbolic comparison tasks (paired, sequential and intermixed presentations) in a sample of 39 adults. They found significant reliability in all tasks, being the Weber fraction values more reliable than those of RT. In general, paired presentation designs (Maloney et al., 2010; Price et al., 2012; Sasanguie et al., 2011) have shown larger internal reliability. This could be explained because these tasks are not influenced by additional cognitive processing demands such as, higher working memory capacity involvement in the sequential condition or visual resolution in intermixed presentations (Price et al., 2012).

In contrast, it has recently been reported by Inglis and Gilmore (2014) that overall accuracy is the most reliable index compared to weber fraction and numerical ratio effect (accuracy and RT) in nonsymbolic comparison tasks. In this study, the NRE was reported to show poor test-retest reliability and no correlations with either Weber fractions or accuracy (Inglis and Gilmore, 2014). Additionally, it has been suggested by Dietrich, Huber, & Nuerk (2015) the interpretation that a smaller NRE/NDE reflects a better ANS acuity is problematic when participants struggle with the task and exhibit performance close to the chance level; since a smaller NRE/NDE might indicate floor effects.

On the other hand, convergent validity studies of the different versions of these tasks (e.g., nonsymbolic vs. symbolic format) have yielded inconsistent results, with some of the studies reporting no correlation between nonsymbolic and symbolic tasks used to explore

ANS characteristics. For example, in the Maloney et al. (2010) study, the authors found no correlation between performance (using numerical distance effect measures) in the nonsymbolic vs. the symbolic version of a small numerosities comparison task. In contrast, data from Gilmore et al. (2011) did show a significant correlation between these tasks (in this study performance was assessed using accuracy). However, they found no correlation between measures when large numerosities were compared.

Additionally, inconsistent evidence is available also concerning whether there is a relation between children individual performance in nonsymbolic and symbolic tasks exploring the ANS precision. Several studies have failed to find correlations between accuracy or numerical distance effect measures in nonsymbolic vs. symbolic comparison tasks involving small numerosities (Desoete et al., 2012; Lonnemann, Linkersdörfer, Hasselhorn, & Lindberg, 2011). Conversely, Holloway and Ansari (2009) found a significant correlation between the mean RT of two (nonsymbolic vs. symbolic) versions of a small numerosities comparison task. Similar results were obtained by Gilmore, Attridge, De Smedt, and Inglis (2014), by correlating the accuracy between the nonsymbolic and symbolic versions of a comparison task involving small numerosities and an approximate addition task. The authors of this study suggest the correlation between performance in nonsymbolic and symbolic comparison tasks in children (but not adults), could be due to differences in numerical skills maturation during development, in that case, children's performance probably could reflect not only the ANS precision, but could also be modeled by interference from domain-general cognitive processes, such as working memory demands necessary to perform the tasks.

Likewise, studies dedicated to explore the relations between basic numerical processing and arithmetic performance have also contributed evidence on the convergent validity of different ANS measures. If both nonsymbolic and symbolic tasks index the ANS, results from both tasks should predict or correlate with arithmetic performance. In this regard the evidence is also inconsistent. Several studies involving children exhibiting typical development of numerical processing or children with arithmetic learning disabilities support in a similar way the relevance of nonsymbolic skills as predictors of subsequent arithmetic performance (Halberda et al., 2012; Libertus et al., 2011; Mussolin et al., 2010; Reigosa-Crespo et al., 2013; Wong, Ho, & Tang, 2015). In contrast, other studies highlight symbolic and mapping skills preponderant role as predictors of arithmetic success in children (Castro et al., 2012; De Smedt & Gilmore, 2011; Landerl & Kölle, 2009; Lonnemann et al., 2011; Rousselle & Noel, 2007; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013).

In the present study, we explore whether the performance in two numerical comparison tasks (nonsymbolic and symbolic) involving small numerosities (4–9) are correlated. Finding a statistically significant correlation between the tasks would indicate they both index the same underlying neurocognitive system, or in other words, that both tasks could be considered as appropriate ANS measures. For this, test-retest reliability and convergent validity analysis between the results of the tasks (overall accuracy, RT and efficiency measures) are conducted. Additionally, the relations between the tasks and arithmetic performance -assessed using a mental arithmetic test- are evaluated. The influence of domain-general cognitive processes (verbal and visuospatial working memory) in numerical processing will be controlled for; since previous studies have reported a significant contribution of these processes to numerical cognition (Alloway & Passolunghi, 2011; Ashkenazi, Rosenberg-Lee, Metcalfe, Swigart, & Menon, 2013; Szucs, Devine, Soltesz, Nobes, & Gabriel, 2013; see Cragg & Gilmore, 2014 for a review).

Methods

Participants

The sample comprised 101 Chilean school children (61 boys); ages ranging between six years and 10 months to 14 years and one month ($M = 9.7$ years). See Table 1 for a detailed sample description by grade and gender. A sample selection criterion of 50th to 95th percentile on the Raven's Colored Progressive Matrices Test (Raven, Court, & Raven, 1992) was used in order to include in the study children exhibiting typical intellectual capacities. Written consent from all parents was obtained, and all participants provided verbal assent for assessments.

Materials

Simple reaction time task

Some children are relatively slow in pressing keys when responding to stimuli. In order to control for that, the RT of all experimental tasks (described below) were

Table 1. Sample Details

School Grade	N	Sex
		Male (female)
Grade 1	15	9 (6)
Grade 2	16	9 (7)
Grade 3	16	10 (6)
Grade 4	20	14 (6)
Grade 5	18	11 (7)
Grade 6	16	8 (8)
Total	101	61 (40)

adjusted taking a Simple RT measure into account (see Statistical Analysis section). Children were presented with a happy yellow face on a white background. This stimulus was counterbalanced, appearing on the left or the right side of the screen. Children were asked to press one of two specific keys, depending on the side where the stimulus appeared. The task consisted of 20 trials. Each trial started with the presentation of the stimulus, which remained on the screen until a response was given. Then, a white screen was presented during a variable (100 to 1500 ms) inter-stimulus interval (ISI). Six practice trials were presented before starting the task.

Numerical tasks

Comparison Tasks: Each task (nonsymbolic and symbolic) consisted of 60 comparison pairs (with numerosities 1 to 9) that varied among four experimental conditions: small ratios (.33 and .50), large ratios (.66, .75 and .85), close numerical distances (1 and 2) and far numerical distances (4 and 5). The trials were presented in two separate blocks of 30 stimuli each. Each trial started with the presentation of a comparison pair until a response was given, followed by an ISI of 500 ms (black screen) during which a red fixation cross remained in full view. Six practice trials were given before starting the task. Comparison pairs for nonsymbolic task consisted of two white squares (side = 55 mm) containing a variable number of white circles. Children were instructed to select the one that contained more elements (or less, according to instruction). Both white squares were presented on a black background and were separated by a red fixation cross (distance between squares = 8 mm). To prevent children to use strategies based on low-level continuous variables, we generated three sets of arrays controlling for density (array density and dot size were kept constant), surface (total occupied area and luminance were kept constant) and area (total occupied area and dot size were kept constant). Comparison pairs for symbolic task consisted of two white Arabic digits (Arial font size 60), presented on a black background. Children were instructed to select the digit with the largest (or smaller, according to instruction) numerical size. Participants were encouraged to answer as quickly as they could without making mistakes.

For this study, only comparison pairs above the subitizing range were analyzed.

Arithmetic Mental Test: Item-timed computerized test consisted of 56 trials presented in two blocks: 28 simple additions and 28 subtractions. All blocks included white Arabic digits, (numerosities 1 to 9) in Arial font size 60, presented on a black background. Items were presented horizontally in the form " $2 + 4$ ".

Below these, two response alternatives, one correct and one incorrect, were simultaneously displayed on the left and right sides of the screen. Incorrect answers were created by adding or subtracting 1 or 2 from the solution. The correct answer position was counterbalanced across trials. Each trial started with the presentation of the stimulus until a response was provided and followed by an ISI of 500 ms (black screen) during which a red fixation cross remained in full view. Six practice trials were given before starting the task. Children were asked to select the correct answer as quickly as they could without making mistakes.

Working memory tasks

Digit span tests are a well-validated measure of working memory thought to involve both the executive and phonological working memory systems. In such tasks, participants listen to a series of numbers and are asked to recite them in order (forward version) or in reverse order (backward version). The Corsi blocks task (Milner, 1971) has been considered as a visuospatial counterpart for the verbal-memory span tasks.

Phonological working memory was assessed using The Digit Span Scale (backward) of the Wechsler Intelligence Scale for Children- Revised (WISC- R) (Sattler, 1982).

Visuospatial working memory: In the present study we used designed a computerized working memory task, similar to the Corsi blocks task. Thus, the visuospatial task was equivalent to the Digit Span test. Children were presented with a grid of 20 squares on a white background. Each trial involved presenting a sequence in which grid squares change color from white to red. Within each sequence, the corresponding stimuli change colors every 300 ms. After the sequence, the screen color changed to pink background and children had to respond by clicking with the mouse the grid squares that changed from white to red (in reverse order the color of squares changed in the original sequence). The task consisted of 14 sequences (2, 3, 4, 5, 6, 7 or 8 stimuli, each numerosity is repeated twice). Four practice trials were given before starting the task. One score was given for each consecutive pair of objects that was remembered in the correct order. In this way, a trial consisting of five objects gave a maximum of four points. The span score was calculated as the sum of scores across 14 trials.

Procedures

Children were individually assessed in a quiet room at the school. The experimental tasks were administered in two sessions of 30 – 40 minutes. In the first session, Raven's Colored Progressive Matrices Test, Working Memory and Simple Reaction Time tasks

were administered. The numerical tasks were administered in the second session: Comparison tasks (in counterbalanced order) and Arithmetic Mental Test.

Statistical analysis

Tasks performance was assessed using accuracy, RT and efficiency measures. Median RT and efficiency measures included correct responses only. All children included in this study obtained a correct responses' percentage above 50% in each experimental condition. The median RT of correct responses per condition were adjusted, subtracting the median of simple RT for each participant out of the corresponding condition median RT (adjRT). This RT adjustment procedure allows controlling for the variability due to individual differences in general processing speed. The efficiency measures (EM) were calculated by dividing adjRT by the proportion of correct responses ($EM = adjRT / \text{proportion of correct responses}$). This is an inverse measure (higher efficiency measures represent worse performance) which seizes the relationship between RT and accuracy (see Table 2 for a detailed global EMs description by grade in numerical comparison tasks and arithmetic achievement).

Shapiro-Wilk's normality test and Levene's variance homogeneity test were employed to test the dependent variables fit parametric assumptions concerning homogeneous variance and normal data distribution. When the variables didn't completely fit the assumptions nonparametric correlations and logarithmic transformations were applied to the variables for the corresponding analysis.

To test for the presence of classical numerical processing effects (NRD and NDE), different repeated measures ANOVAs for both presentation formats (nonsymbolic and symbolic) were conducted on adjRT data. Previous works focused on the assessment of numerical mental representations via the ratio and distance effects have used different formulas to calculate these effects. This may as well be a reason for the previously described inconsistency among the studies. Thus, in order to

avoid the effect of choosing a specific formula, over other possible ones, the results of each task were collapsed into four experimental conditions: small (.33 and .50) and large numerical ratios (.66, .75 and .85), close (1 and 2) and far numerical distances (4 and 5).

Reliability and convergent validity using the corresponding accuracies, EMs or median of adjRT analyses were conducted. Test-retest analysis was conducted to assess reliability. Accuracies, efficiency and adjRT measures for the first block of trials corresponding to each experimental condition and global block measures were correlated with the corresponding measures of the second block of trials (for nonsymbolic and symbolic tasks). To assess convergent validity between tasks, accuracy, adjRT and efficiency measures for each experimental condition and global task measures of the nonsymbolic task were correlated with the corresponding experimental condition measures in the symbolic task.

In order to account for the significant covariations between domain-general cognitive processes and numerical processing systematically reported in the literature, additional convergent validity analysis controlling for the effects of domain-general cognitive processes (verbal and visuospatial working memory) and age were conducted via partial correlations and factor analysis. Likewise, to explore the relationship between the ANS measures and arithmetic achievement, partial correlations controlling for the effect of working memory processes and age between global nonsymbolic and symbolic efficiency measures and global efficiency in mental arithmetic were conducted.

On the other hand, to explore developmental trends in the data, correlations between global efficiency measures and age were conducted. Finally, to explore whether the relationships among tasks varied with age, a correlation between age and the residuals resulting from regressing the symbolic global efficiency onto the nonsymbolic global efficiency was conducted, following the procedure previously used by Gilmore et al., 2014. Additionally, an ANOVA was conducted including the residuals of the regression of EM in comparison tasks (symbolic vs. nonsymbolic) as dependent variable and school grades as between factor (grades 1 to 6) in order to describe the possible interactions between the relationship symbolic and non symbolic numeric comparison by school grade.

Results

Numerical Effects Analyses: Ratio and Distance Effects

In order to determine whether the numerical processing effects (NRE and NDE) were elicited by the tasks, data was analyzed using two repeated measures ANOVAs with experimental conditions as within-subject factors:

Table 2. Mean of Global Efficiency Measures (EMs) by Grade for All Tasks

School Grade	N	Arithmetic task	Nonsymbolic comparison task	Symbolic comparison task
Grade 1	15	7,528.59	1,630.84	1,604.91
Grade 2	16	5,613.86	1,448.48	1,297.83
Grade 3	16	3,766.79	1,274.53	1,173.16
Grade 4	20	2,622.28	1,159.46	967.45
Grade 5	18	2,943.23	1,038.42	942.87
Grade 6	16	2,344.74	1,161.94	875.57

1. Numerical ratio (small and large) as within-subject factor, and 2. Numerical distance (close and far) as within-subject factor. These analyses revealed both tasks elicited the classical numerical effects in the group: Nonsymbolic task: NRE, $F(1, 100) = 262.39$, $MSE = .026$, $p < .001$; and NDE, $F(1, 100) = 234.69$, $MSE = .031$, $p < .001$. Symbolic task: NRE, $F(1, 100) = 260.47$, $MSE = .003$, $p < .001$; and NDE $F(1, 100) = 329.50$, $MSE = .003$, $p < .001$.

Reliability Analysis

Reliability analyses were performed using accuracy, median of adjRT and EMs calculated for each experimental condition: small and large ratios, and close and far numerical distances. Additionally, reliability was explored for each task using overall accuracy, adjRT and EM. Significant correlations ($p < .001$) were found between Block 1 and 2 for each experimental condition in nonsymbolic and symbolic tasks (for both adjRT and EMs). The analysis between global adjRT and global EMs for the first and second blocks of trials showed significant correlation too, for both nonsymbolic and symbolic tasks. Low or no statistically significant correlations between the accuracy measures for Blocks 1 and 2 were found. See details in Table 3 and Figures 1 and 2.

Convergent Validity Analysis. Correlations between nonsymbolic and symbolic tasks

To analyze convergent validity the corresponding accuracy, adjRT and EMs for each experimental condition between tasks were correlated. Significant correlations ($p < .001$) were found between the experimental conditions of both tasks (for both adjRT and EMs). The global tasks analysis between nonsymbolic and symbolic comparison tasks showed significant correlation too, for both adjRT and EMs. In all cases, accuracy showed lower or non-statistically significant correlations. See details in Table 4 and Figures 3 and 4.

Numerical comparison and working memory tasks' correlation analysis

In order to analyze the interaction, and possible contribution of domain-general cognitive processes to performance in numerical comparison, a correlation analysis including nonsymbolic and symbolic global EMs and visuospatial and verbal working memory measures was conducted.

Significant negative correlations were found between global EM in nonsymbolic comparison task and both, backwards visuospatial working memory scores ($r = -.34$, $p < .01$) and backwards digit span scores ($r = -.24$, $p < .05$); and between symbolic global EM and backwards visuospatial working memory scores ($r = -.59$, $p < .001$) and backwards digit span scores ($r = -.35$, $p < .001$).

Likewise, significant negative correlations between adjRT in nonsymbolic comparison task and both, backwards visuospatial working memory scores ($r = -.32$, $p < .01$) and backwards digit span scores ($r = -.21$, $p < .05$) were found. In addition, significant negative correlations between adjRT in symbolic comparison task and both, backwards visuospatial working memory scores ($r = -.56$, $p < .001$) and backwards digit span scores ($r = -.23$, $p < .01$) were found.

No significant correlations between accuracy in comparison tasks and backwards visuospatial and verbal working memory scores were found.

Analysis controlling for the effect of working memory processes

The partial correlation analyses between global nonsymbolic and symbolic EMs and adjRT were conducted, controlling for the contribution of available verbal and visuospatial (backward) working memory measures and age, to numerical comparison. Again, high and statistically significant correlation ($p < .001 - p < .05$) supported the convergent validity between nonsymbolic and symbolic tasks (See Table 4 for details).

Table 3. Correlations between Blocks 1 and 2 (adjRT, EMs and Accuracy) for Each Experimental Condition and Global Task Performance in the Nonsymbolic and Symbolic Tasks

Experimental condition	Correlation coefficient (r)					
	adjRT Nonsymbolic task	adjRT Symbolic task	EM Nonsymbolic task	EM Symbolic task	Accuracy Nonsymbolic task	Accuracy Symbolic task
Global task	.78***	.86***	.74***	.86***	.30**	.18
Small ratio	.81***	.72***	.74***	.69***	.05	.08
Large ratio	.78***	.82***	.71***	.78***	.33**	.21*
Close distance	.66***	.81***	.68***	.71***	.24	.16
Far distance	.83***	.79***	.68***	.73***	.06	.27*

Note: * $p < .05$; ** $p < .01$; *** $p < .001$

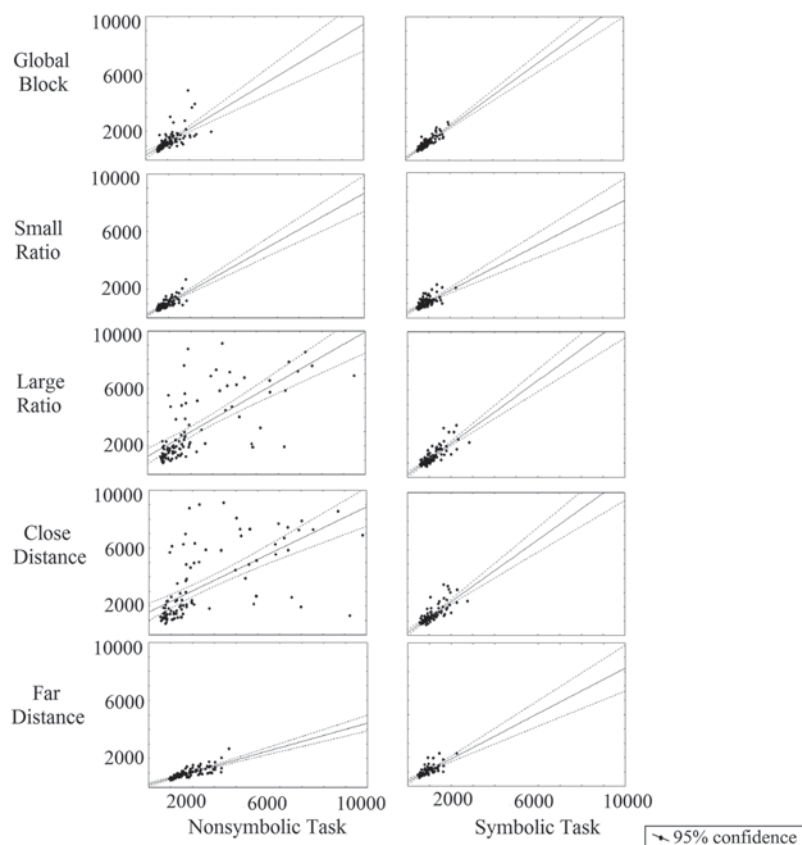


Figure 1. Median of adjusted RT (ms) for block 1 vs. block 2 in the nonsymbolic and symbolic comparison tasks. Horizontal axis: adjusted RT (AdjRT) for block 1; Vertical axis: AdjRT for block 2. The top row shows scatterplots of AdjRT for global block. On the second and third rows, scatterplots of AdjRT for small and large numerical ratio conditions are presented. On the two last rows scatterplots of AdjRT for close and far numerical distance conditions are presented.

Factor analysis

Principal components analyses with varimax rotation were conducted in order to examine clustering among variables. A cutoff point of $\pm .70$ was used on rotated factor loadings for principal factors extraction to include a variable in a factor. The resulting two factor solution accounted for 74.77 % of the cumulative variance, indicating that these solutions effectively characterized the variance in these data. The first factor included the global nonsymbolic and symbolic numerical comparison EMs (domain-specific factor, factor loadings: .93 and .79 respectively). The second factor included the backward visuospatial span and the backward Digit Span scores (domain-general factor loadings: .84 and .72 respectively).

Correlation between ANS and Exact Mental Arithmetic

Partial correlations between global nonsymbolic and symbolic EMs (controlling for the effect of domain-general cognitive processes) and global EM in mental arithmetic showed high and statistically significant correlations between exact mental arithmetic and both,

nonsymbolic ($r = -.29, p < .01$) and symbolic efficiency ($r = -.46, p < .001$).

Developmental trends' analysis

Significant negative correlation between global efficiency measures and age were found in both, nonsymbolic ($r = -.29, p < .01$) and symbolic ($r = -.62, p < .001$) comparison tasks. Likewise, significant negative correlation between age and the residuals resulting from regressing the symbolic global efficiency onto the nonsymbolic global were found ($r = -.55, p < .001$), suggesting the relationships among the tasks varied across the samples' age range. The ANOVA conducted on the residuals of the regression of efficiency in symbolic vs. nonsymbolic comparison tasks as dependent variable, and the school grade as the between factor showed a significant effect of grade, $F(5, 87) = 7.8084, p < .001$. Planned comparison analysis performed showed there was no statistically significant difference in the relationship between the tasks for children in grade 1 and grade 2, but that both groups are significantly different from children of all the remaining grades. Likewise, second graders are not significantly different from

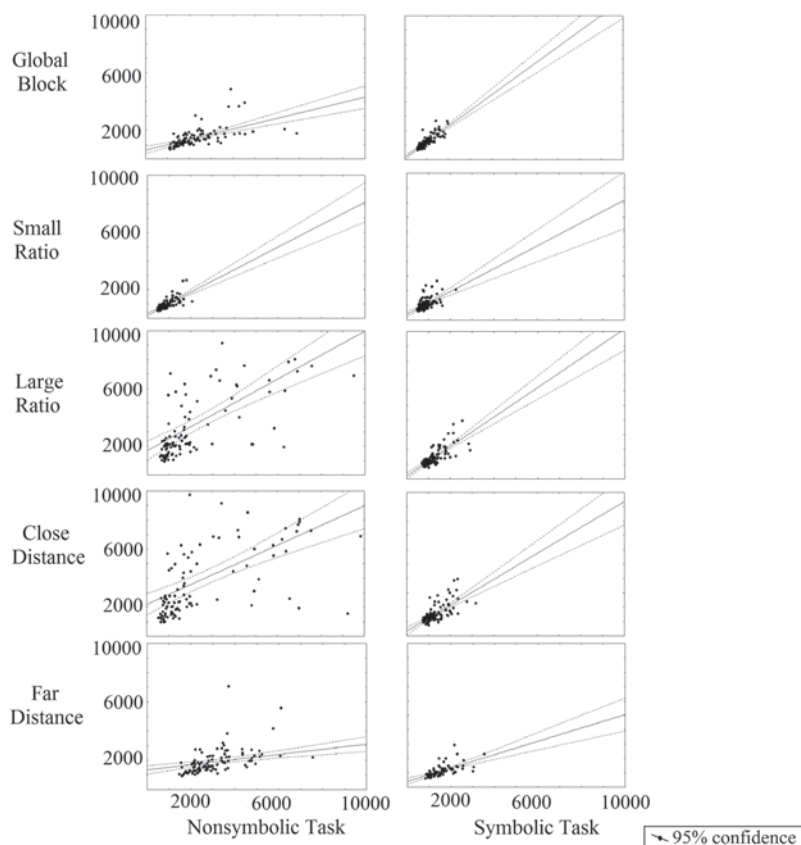


Figure 2. Efficiency measures (EM) for block 1 vs. block 2 in the nonsymbolic and symbolic comparison tasks. Horizontal axis: EM for block 1; Vertical axis: EM for block 2. The top row shows scatterplots of EM for global block. On the second and third rows, scatterplots of EM for small and large numerical ratio conditions are presented. On the two last rows scatterplots of EM for close and far numerical distance conditions are presented.

Table 4. Correlations between nonsymbolic and symbolic tasks (adjRT and EM) for each experimental condition and global task

Experimental condition	Correlation coefficient (r)		
	adjRT	EMs	Accuracy
Global task	.71*** (.69***)	.70 *** (.66***)	.34*** (.32**)
Small ratio	.76 *** (.73***)	.75 *** (.72***)	.11 (.14)
Large ratio	.56*** (.59***)	.54 *** (.56***)	.33 *** (.31**)
Close distance	.51 ** (.55***)	.46 ** (.47***)	.28 ** (.26**)
Far distance	.77 *** (.75***)	.75 *** (.72***)	.18 (.21*)

Note: Partial correlations controlling for age, visual and verbal working memory, are shown in parentheses.
 * $p < .05$; ** $p < .01$; *** $p < .001$.

third graders, but are significantly different compared to fourth, fifth and sixth graders. In contrast, third graders are only significantly different from sixth graders. No statistically significant differences among fourth, fifth and sixth graders were found (see Table 5 and figure 5).

Discussion

In the present study, we report adequate reliability and convergent validity for two comparison tasks

(nonsymbolic and symbolic) involving small numerosities, designed for the ANS assessment. The correlations found between the tasks both in efficiency measures and adjRT suggest the same neurocognitive system underlies nonsymbolic and symbolic tasks. Also, the performance on mental arithmetic tasks significantly covaried with the ANS efficiency measures. Finally, a developmental trends analysis performed on the global efficiency measures and grade showed

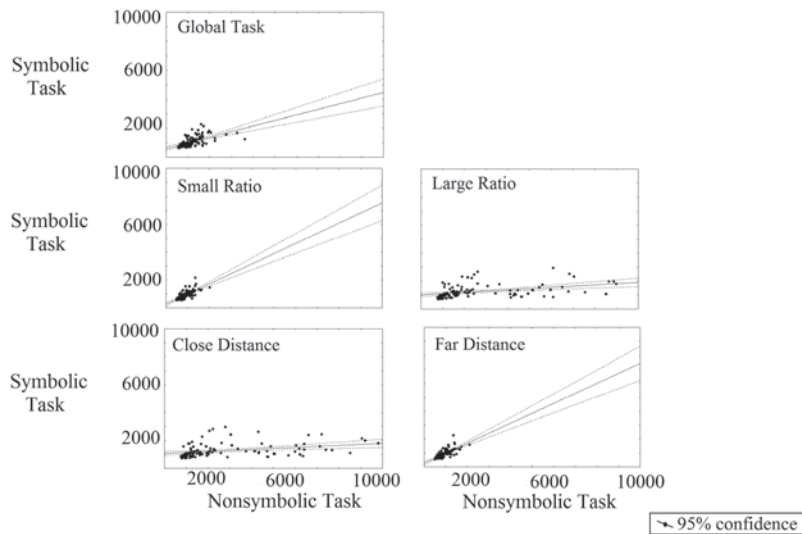


Figure 3. Median of adjusted RT (ms) for nonsymbolic vs. symbolic comparison tasks. Horizontal axis: adjusted RT (AdjRT) for nonsymbolic task; Vertical axis: AdjRT for symbolic tasks. The top row shows the scatterplot of AdjRT for global task. On the second row scatterplots of AdjRT for small and large numerical ratio are presented. On the last row scatterplots of AdjRT for close and far numerical distance conditions are presented.

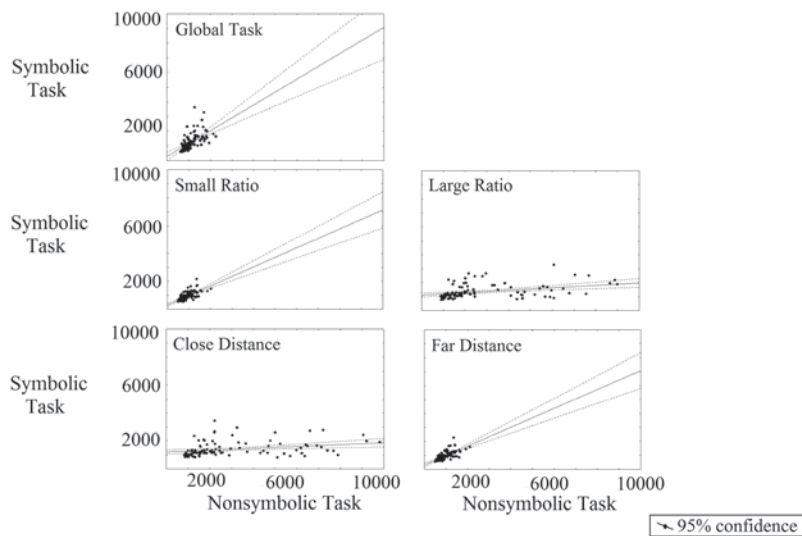


Figure 4. Efficiency measures (EM) for nonsymbolic vs. symbolic comparison tasks. Horizontal axis: EM for nonsymbolic task; Vertical axis: EM for symbolic tasks. The top row shows the scatterplot of EM for global task. On the second row scatterplots of EM for small and large numerical ratio are presented. On the last row scatterplots of EM for close and far numerical distance conditions are presented.

Table 5. Differences between School Grades in the Residuals of EMs for Symbolic vs. Nonsymbolic Comparison Tasks

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Grade 2	–				
Grade 3	2.86 **	–			
Grade 4	4.66 ***	2.72 **	–		
Grade 5	4.17 ***	2.31 *	–	–	
Grade 6	5.39 ***	3.55 ***	2.47 *	–	–

Note: * $p < .05$; ** $p < .01$; *** $p < .001$; – (not significant)

the relationship between nonsymbolic and symbolic numeric comparison varies with age. Surprisingly, all analysis including accuracies showed low or non-statistically significant figures.

Regarding the reliability analysis, our data is consistent with previous studies (Gilmore et al., 2011; 2014; Maloney et al., 2010; Price et al., 2012; Sasanguie et al., 2011). The results showed both high and significant correlations among individual adjusted RT and efficiency measures between the two blocks of stimuli in both comparison tasks (symbolic and nonsymbolic).

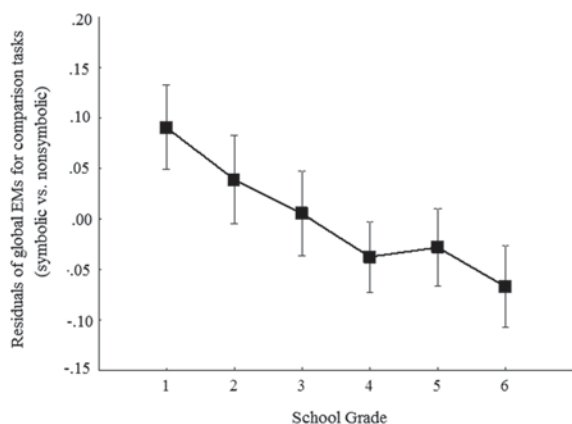


Figure 5. Differences in the residuals of EMs for symbolic vs. nonsymbolic comparison tasks by grade.

Thus, in both tasks, measures calculated using data from the first half of the experiment are highly predictive of the subjects' results in the second half of the experiment, increasing the probability of detecting differences between groups. These results support the test-retest reliability in both tasks (nonsymbolic and symbolic). The correlations involving adjusted RT and efficiency measures were above the Cohen and Swerdlik's (2009) acceptable range of reliability ($r > .65$). In contrast, overall accuracy and accuracy per experimental conditions were below this range of reliability. Probably, these results reflect a ceiling effect, since all the children in the sample exhibited accuracies significantly above chance level in all experimental conditions. Nevertheless, taking into consideration recent reports on the favorable statistical properties of overall accuracy as an index of ANS acuity (Inglis & Gilmore, 2014); overall efficiency measures, which reflect the reaction time/accuracy trade off, were included as dependent variables in the rest of the analysis.

Concerning the convergent validity analysis, the global task descriptors used (global individual adjusted RT and efficiency measures) exhibited high and significant correlations between the nonsymbolic and symbolic tasks. Furthermore, the data analysis conducted by experimental condition, showed high and statistically significant correlation values across all experimental conditions. In contrast, previous studies using numerical distance effect measures have failed to find significant correlation between nonsymbolic and symbolic numerical comparison tasks (Gilmore et al., 2011; Maloney et al., 2010). Likewise, others have found no correlations between different measures within the same task (e.g., between ratio effects and accuracy in Inglis & Gilmore, 2014; between Weber fraction estimates and ratio effects in Price et al., 2012). Such results support the idea that most of the differences between previous studies' reports concerning reliability and

convergent validity between the nonsymbolic and symbolic versions of comparison tasks may primarily be due to the use of different behavioral measures. Note the ratio and distance effects successfully indexed participants' performance variability within certain parameters (e.g., close distance vs. far distance) but were not able to capture the absolute task performance levels as achieved for example, with efficiency measures (combining both RT and overall accuracy of the tasks). On the other hand, we consider the way we explored the ANS using performance per condition (using efficiency measures) and the global tasks performance allows to avoid the effect of choosing a specific formula to calculate the numerical processing effects and could be employed for future reliability and validity studies, as well as for other studies describing numerical cognition, in order to contribute a more homogenous and comparable performance reference framework. Additionally, the detailed convergent validity analysis conducted on the adjusted RT and efficiency measures for all (corresponding) experimental conditions between the nonsymbolic and symbolic numerical tasks used in the present study provides a robust description on the extent to which the tasks similarly index the ANS, not usually available in convergent validity analysis. Note again, accuracy yielded low or non-statistically significant convergent validity values. This result is consistent (or may be a result) of the aforementioned lower reliability found for this measure.

In general, our findings support convergent validity between nonsymbolic and symbolic comparison tasks, as ANS assessment measures. These results concur with those obtained in children by Gilmore et al. (2014), and contrast with previous research in adults (Gilmore et al., 2011; Maloney et al., 2010). Gilmore et al. (2014) hypothesize that differences in convergent validity between studies in children and adults may be due to developmental differences in domain-general cognitive processes whose influence may interact differently with the ANS performance for nonsymbolic and symbolic tasks. In other words, there may be differences between children and adults in the extent to which performance on these tasks reflects ANS acuity vs. domain-general cognitive demands.

Several studies have highlighted the involvement of domain-general cognitive processes (inhibitory control, working memory, executive control) on numerical processing during childhood (Alloway & Passolunghi, 2011; Ashkenazi et al., 2013; Clayton & Gilmore, 2015; Cragg & Gilmore, 2014; Fuhs & McNeil, 2013; Gilmore et al., 2013; Szucs et al., 2013). For this reason, in order to control for verbal and visuospatial working memory contributions to numerical processing, additional convergent validity analysis between global efficiency

measures for nonsymbolic and symbolic comparison tasks were conducted. The results showed high and statistically significant correlations between the corresponding global efficiency measures even after controlling for the effect of these domain-general cognitive processes. Concurrent evidence on the specific contribution of ANS to numerical performance was obtained via Factor Analysis, using a Principal Components Analysis method for factors extraction. This analysis showed a two factor solution, the first one of them fundamentally explaining the variance of both numerical comparison tasks and the second one primarily explaining the variance of the two (backward) verbal and visuospatial working memory tasks administered to the children. These results support the convergent validity between the nonsymbolic and symbolic numerical comparison tasks and suggest that performance on both tasks may be driven by the same underlying processing; or in other words, the results suggest both tasks index the ANS. Additionally, these results reflect the differences in terms of cognitive architecture concerning specific numerical abilities and domain-general cognitive processes. The factor solution found here contrasts with evidence reported in adults, showing very low correlations in similar tasks (Gilmore et al., 2011; Maloney et al., 2010) and challenges the hypothesis posed by previous studies explaining the differences in performance between children and adults in terms of the impact of developmental differences in domain-general cognitive process required by differently demanding experimental tasks (e.g., Gilmore et al., 2014).

Nevertheless, the factor analysis results could be grouping the two comparison tasks together due to the fact they both have a comparative element to them and require similar domain-general processing, in contrast to the different methodological and response format of the working memory tasks. A more rigorous way of dealing with this would be to include additional tasks, already described in the literature as reliable indices of the ANS in the factor analysis, in order to test whether numerical comparison tasks and this additional indices of the ANS are grouped together, supporting the existence of distinct cognitive architectures underlying the tasks, or whether the current results reflect only format characteristics of the tasks.

On the other hand, the present study supports the relevance of the ANS in the typical development of arithmetic abilities during childhood. Note the statistically significant and high correlations found between nonsymbolic and symbolic numerical comparison tasks and mental arithmetic remain, despite the contribution of domain-general cognitive processes to performance was controlled for. These results are consistent with previous reports on children with low math achievement exhibiting significant differences in basic numerical

processing when compared to controls, even when controlling for working memory processes (Castro et al., 2012; Wong et al., 2015). Nevertheless it is worth mentioning the analysis controlling for the effect of verbal and visuospatial working memory processes in the sample presented here contributes new relevant evidence in typically developing children, not available in previous reliability/convergent validity studies, to the best of our knowledge. On this regard, other domain-general cognitive processes (not assessed in the present sample), have also been described to impact on arithmetic achievement and basic numerical abilities. That is the case of attention (LeFevre et al., 2013; Swanson, 2011), inhibitory control (Clayton & Gilmore, 2015; Cragg & Gilmore, 2014; Fuhs & McNeil, 2013; Gilmore et al., 2013; Szucs et al., 2013), intellectual capacity (Alloway & Passolunghi, 2011), and phonological awareness (De Smedt, Taylor, Archibald, & Ansari, 2010; Krajewski & Schneider, 2009), just to mention some of them. Future studies could also consider controlling for these variables in the analysis.

The developmental trends' analysis conducted showed global efficiency in both nonsymbolic and symbolic tasks significantly improve with age. Consistent results describing increases in numerical comparison efficiency and in the acuity of the ANS have been previously reported by cross-sectional (Castro et al., 2011; Landerl & Kölle, 2009) and longitudinal studies (Libertus et al., 2011; Reigosa et al., 2013; Sasanguie et al., 2013). Additionally, the developmental trends' analysis showed the relationship between nonsymbolic and symbolic numerical comparison also varies across the sample's age range. Though it is not possible using the analysis performed in the present study to provide deeper insight on its nature or characteristics, the variation in the relationship between the nonsymbolic and symbolic overall efficiency described here, showing decreasing differences between the tasks with age, may reflect the effect of experience in the ANS acuity, or; on a more basic neurocognitive level this could be indicative of the development of the interphase between the auditory-verbal, the analogue and the visual-digital representational codes for numbers becoming more accessible and automatic across developmental time. The ANOVA showed how with age, the prediction concerning performance in the symbolic comparison task, calculated on the basis of performance in the nonsymbolic comparison task is increasingly accurate. This suggests that from third grade on, children access to symbolic numeric representations from available nonsymbolic information becomes more automatic and also suggests increasingly higher levels of development of the nonsymbolic to symbolic interphase. Future studies, preferably longitudinal, or including a larger age span, from infants to adults, and which take into account the

effects of other cognitive processes in numerical cognition are required in order to clarify the present results.

Finally, it's important to note the results described in the present study were obtained using correlation analyses, which allow to describe covariation patterns between variables; but are not designed to disentangle causality relations between these variables (is the variability in one of variables causing the values in the other one to vary; or is there a third variable, or combination of variables, underlying the variability of them both?). In addition to the partial correlations analysis employed here, other statistical analysis techniques should be considered in future studies in order to better describe the interactions and the directionality of the interactions between the relevant variables. Furthermore, the theoretical interpretations of the covariation patterns between nonsymbolic and symbolic measures as indices of the ANS adopted here relied on a series of previous studies; but no additional ANS measures, other than the two numerical comparison tasks included in the convergent validity analysis performed, were included in the present study. Additional measures, which could act as golden rule criteria (valid or independent indices) of the ANS could also be included in future studies in order to test for the construct validity of these tasks, for example: the property of scalar variability (Whalen et al., 1999): The mean responses and standard deviation of the responses in an estimation task are actually proportional to each other as the numerosity to be estimated varies, hence the ratio between the standard deviation and the mean response is constant across all numerosities.

In general, the results presented here support both the reliability and convergent validity between nonsymbolic and symbolic comparison tasks and its use to the behavioral assessment of the ANS mental representations. Additionally, according to the evidence described here, the ANS is informative of arithmetic efficiency variability during childhood. The results also show these tasks are sensible to developmental variations in the ANS captured by the tasks overall efficiency in this sample's age range, probably concerning the extent to which efficiency in nonsymbolic and symbolic tasks reflect ANS acuity and/or the development of the analogue-symbolic interphase involved in number processing.

References

- Alloway T. P., & Passolunghi M. C. (2011). The relationship between working memory, IQ, and mathematical skills in children. *Learning and Individual Differences*, 21(1), 133–137. <https://doi.org/10.1016/j.lindif.2010.09.013>
- American Educational Research Association, American Psychological Association y National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Ashkenazi S., Rosenberg-Lee M., Metcalfe A. W. S., Swigart A. G., & Menon V. (2013). Visuo-spatial working memory is an important source of domain-general vulnerability in the development of arithmetic cognition. *Neuropsychologia*, 51(11), 2305–2317. <https://doi.org/10.1016/j.neuropsychologia.2013.06.031>
- Castro D., Estévez N., & Pérez O. (2011). Typical development of quantity comparison in school-aged children. *The Spanish Journal of Psychology*, 14(1), 50–61. https://doi.org/10.5209/rev_SJOP.2011.v14.n1.4
- Castro D., Reigosa V., & González E. (2012). Non-symbolic and symbolic number magnitude processing in children with developmental dyscalculia. *The Spanish Journal of Psychology*, 15(3), 952–966. https://doi.org/10.5209/rev_SJOP.2012.v15.n3.39387
- Clayton S., & Gilmore C. (2015). Inhibition in dot comparison tasks. *ZDM: Mathematics Education*, 47, 759–770. <https://doi.org/10.1007/s11858-014-0655-2>
- Cohen R. J., & Swerdlik M. (2009). *Psychological testing and assessment: An introduction to tests and measurement* (7th ed.). New York, NY: McGraw-Hill.
- Cragg L., & Gilmore C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, 3(2), 63–68. <https://doi.org/10.1016/j.tine.2013.12.001>
- 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, 278–292. <https://doi.org/10.1016/j.jecp.2010.09.003>
- De Smedt B., Taylor J., Archibald L., & Ansari D. (2010). How is phonological processing related to individual differences in children's arithmetic skills. *Developmental Sciences*, 13, 508–520. <https://doi.org/10.1111/j.1467-7687.2009.00897.x>
- Dehaene S., & Changeux J. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neurosciences*, 5, 390–407. <https://doi.org/10.1162/jocn.1993.5.4.390>
- Desoete A., Ceulemans A., De Weerd F., & Pieters S. (2012). Can we predict mathematical learning disabilities from symbolic and non-symbolic comparison tasks in kindergarten? Findings from a longitudinal study. *British Journal of Educational Psychology*, 82, 64–81. <https://doi.org/10.1348/2044-8279.002002>
- Dietrich J. F., Huber S., & Nuerk H. C. (2015). Methodological aspects to be considered when measuring the approximate number system (ANS) – a research review. *Frontiers in Psychology*, 6, 295. <https://doi.org/10.3389/fpsyg.2015.00295>
- Fuhs M. W., & McNeil N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: Contributions of inhibitory control. *Developmental Science*, 16(1), 136–148. <https://doi.org/10.1111/desc.12013>
- Gilmore C., Attridge N., De Smedt B., & Inglis M. (2014). Measuring the approximate number system in children: exploring the relationships among different tasks. *Learning*

- and Individual Differences, 29, 50–58. <https://doi.org/10.1016/j.lindif.2013.10.004>
- Gilmore C., Attridge N., Clayton S., Cragg L., Johnson S., Marlow N., ... Inglis M.** (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS ONE* 8(6), e67374. <https://doi.org/10.1371/journal.pone.0067374>
- Gilmore C., Attridge N., & Inglis M.** (2011). Measuring the approximate number system. *The Quarterly Journal of Experimental Psychology*, 64(11), 2099–2109. <https://doi.org/10.1080/17470218.2011.574710>
- 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>
- Halberda J., Ly R., Wilmer J. B., Naiman D. Q., & Germine L.** (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences*, 109, 11116–11120. <https://doi.org/10.1073/pnas.1200196109>
- Halberda J., Mazocco M. M. M., & Feigenson L.** (2008). Individual differences in nonverbal number acuity correlate with maths achievement. *Nature*, 455, 665–668. <https://doi.org/10.1038/nature07246>
- 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, 17–29. <https://doi.org/10.1016/j.jecp.2008.04.001>
- Inglis M., & Gilmore C.** (2014). Indexing the approximate number system. *Acta Psychologica*, 145, 147–155. <https://doi.org/10.1016/j.actpsy.2013.11.009>
- Iuculano T., Tang J., Hall C. W. B., & Butterworth B.** (2008). Core information processing deficits in Developmental Dyscalculia and low numeracy. *Developmental Science*, 11(5), 669–680. <https://doi.org/10.1111/j.1467-7687.2008.00716.x>
- Izard V., & Dehaene S.** (2007). Calibrating the mental number line. *Cognition*, 106(3), 1221–1247. <https://doi.org/10.1016/j.cognition.2007.06.004>
- Krajewski K., & Schneider W.** (2009). Exploring the impact of phonological awareness, visual-spatial working memory, and preschool quantity-number competencies on mathematics achievement in elementary school: Findings from a 3-year longitudinal study. *Journal of Experimental Child Psychology*, 103, 516–531. <https://doi.org/10.1016/j.jecp.2009.03.009>
- 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>
- LeFevre J., Berrigan L., Vendetti C., Kamawar D., Bisanz J., Skwarchuk S., & Smith-Chant B.** (2013). The role of executive attention in the acquisition of mathematical skills for children in grades 2 through 4. *Journal of Experimental Child Psychology*, 114(2), 243–261. <https://doi.org/10.1016/j.jecp.2012.10.005>
- Libertus M. E., Feigenson L., & Halberda J.** (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science*, 14, 1292–1300. <https://doi.org/10.1111/j.1467-7687.2011.01080.x>
- Lonnemann J., Linkersdörfer J., Hasselhorn M., & Lindberg S.** (2011). Symbolic and non-symbolic distance effects in children and their connection with arithmetic skills. *Journal of Neurolinguistics*, 24, 583–591.
- Maloney E. A., Risko E. F., Preston F., Ansari D., & Fugelsang J.** (2010). Challenging the reliability and validity of cognitive measures: The case of the numerical distance effect. *Acta Psychologica*, 134, 154–161. <https://doi.org/10.1016/j.actpsy.2010.01.006>
- Mazzocco M. M. M., Feigenson L., & Halberda J.** (2011). Preschoolers’ precision of the approximate number system predicts later school mathematics performance. *PLoS ONE*, 6, e23749. <https://doi.org/10.1371/journal.pone.0023749>
- Milner B.** (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin*, 27, 272–277. <https://doi.org/10.1093/oxfordjournals.bmb.a070866>
- Mussolin C., Mejias S., & Noël M.-P.** (2010). Symbolic and nonsymbolic number comparison in children with and without Dyscalculia. *Cognition*, 115, 10–25. <https://doi.org/10.1016/j.cognition.2009.10.006>
- Price G. R., Palmer D., Battista C., & Ansari D.** (2012). Nonsymbolic numerical magnitude comparison: Reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychologica*, 140, 50–57. <https://doi.org/10.1016/j.actpsy.2012.02.008>
- Raven J. C., Court J., & Raven J.** (1992). *Manual for Raven’s progressive matrices and vocabulary scales*. Oxford, UK: Oxford Psychologists Press.
- Reigosa-Crespo V., González-Alemañy E., León T., Torres R., Mosquera R., & Valdés-Sosa M.** (2013). Numerical capacities as domain-specific predictors beyond early mathematics learning: A longitudinal study. *PLoS ONE*, 8(11), e79711. <https://doi.org/10.1371/journal.pone.0079711>
- 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>
- Sasanguie D., Defever E., van den Bussche E., & Reynvoet B.** (2011). The reliability of and the relation between non-symbolic numerical distance effects in comparison, same-different judgments and priming. *Acta Psychologica*, 136, 73–80. <https://doi.org/10.1016/j.actpsy.2010.10.004>
- 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, 418–431. <https://doi.org/10.1016/j.jecp.2012.10.012>
- Sattler J.** (1982). *Assessment of children’s intelligence and special abilities*. Boston, MA: Allyn & Bacon.
- Swanson H. L.** (2011). Working memory, attention, and mathematical problem solving: A longitudinal study of

- elementary school children. *Journal of Educational Psychology*, 103(4), 821–837. <https://doi.org/10.1037/a0025114>
- Szucs D., Devine A., Soltesz F., Nobes A., & Gabriel F.** (2013). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex*, 49(10), 2674–2688. <https://doi.org/10.1016/j.cortex.2013.06.007>
- Waechter S., Stolz J. A., & Besner D.** (2010). Visual word recognition: On the reliability of repetition priming. *Visual Cognition*, 18(4), 537–558. <https://doi.org/10.1080/13506280902868603>
- Whalen J., Gallistel C. R., & Gelman R.** (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, 10, 130–137. <https://doi.org/10.1111/1467-9280.00120>
- Wong T. T. Y., Ho C. S. H., & Tang J.** (2017). Defective number sense or impaired access? Differential impairments in different subgroups of children with mathematics difficulties. *Journal of learning disabilities*, 50(1), 49–61. <https://doi.org/10.1177/0022219415588851>