

Empirical Research

The Relationship Between Children's Approximate Number Certainty and Symbolic Mathematics

Carolyn Baer*^a, Darko Odic^a^[a] Department of Psychology, University of British Columbia, Vancouver, Canada.

Abstract

Why do some children excel in mathematics while others struggle? A large body of work has shown positive correlations between children's Approximate Number System (ANS) and school-taught symbolic mathematical skills, but the mechanism explaining this link remains unknown. One potential mediator of this relationship might be children's numerical metacognition: children's ability to evaluate how sure or unsure they are in understanding and manipulating numbers. While previous work has shown that children's math abilities are uniquely predicted by symbolic numerical metacognition, we focus on the extent to which children's non-symbolic/ANS numerical metacognition, in particular sensitivity to certainty, might be predictive of math ability, and might mediate the relationship between the ANS and symbolic math. A total of 72 children aged 4–6 years completed measures of ANS precision, ANS metacognition sensitivity, and the Test of Early Mathematical Ability (TEMA-3). Our results replicate many established findings in the literature, including the correlation between ANS precision and the TEMA-3, particularly on the Informal subtype questions. However, we did not find that ANS metacognition sensitivity was related to TEMA-3 performance, nor that it mediated the relationship between the ANS and the TEMA-3. These findings suggest either that metacognitive calibration may play a larger role than metacognitive sensitivity, or that metacognitive differences in the non-symbolic number perception do not robustly contribute to symbolic math performance.

Keywords: number sense, symbolic mathematics, certainty, metacognition

Journal of Numerical Cognition, 2020, Vol. 6(1), 50–65, <https://doi.org/10.5964/jnc.v6i1.220>

Received: 2019-05-03. Accepted: 2019-09-26. Published (VoR): 2020-06-15.

*Corresponding author at: Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, British Columbia, Canada V6T 1Z4.
E-mail: cebaer@psych.ubc.ca



This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License, CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Why do some children struggle with learning formal mathematics while others excel with ease? Recent work in cognitive development has focused on the role of children's earliest intuitions about number in predicting individual differences in school-taught symbolic mathematics (Chen & Li, 2014; Dehaene, 2009; Feigenson, Libertus, & Halberda, 2013; Halberda, Mazocco, & Feigenson, 2008; Schneider et al., 2017; Szklarek & Brannon, 2017; but see Gilmore et al., 2013; Szűcs & Myers, 2017). These intuitions—most often tied to the Approximate Number System (ANS)—are fundamentally imprecise, resulting in difficulty when perceptually discriminating small number ratios (e.g., 20 vs. 19 dots, as in Figure 1) versus large number ratios (e.g., 20 vs. 10 dots), but are open found in human newborns, preschoolers, and adults, and even in many non-human animals (for reviews, see Feigenson, Dehaene, & Spelke, 2004; Odic & Starr, 2018; Vallortigara, 2017). ANS representations even allow for simple arithmetic operations, such as addition, subtraction, and division, to be carried over them (Gilmore, McCarthy, & Spelke, 2007; McCrink & Spelke, 2016; Szklarek & Brannon, 2018).

An emerging body of work suggests that individual differences in ANS precision—the smallest ratio that observers can reliably discriminate without counting—predict scores on standardized and non-standardized math assessments in both children and adults (Chen & Li, 2014; Dehaene, 2009; Feigenson et al., 2013; Halberda et al., 2008; Schneider et al., 2017; Szklarek & Brannon, 2017). While these results have frequently been a topic of fierce debate (e.g., Gilmore et al., 2013; Lindskog & Winman, 2016; Sasanguie, Defever, Maertens, & Reynvoet, 2014; Szűcs & Myers, 2017), ANS precision has been shown to predict mathematics ability concurrently (Libertus, Feigenson, & Halberda, 2011; Libertus, Odic, & Halberda, 2012), retrospectively (Halberda et al., 2008), and predictively (Starr, Libertus, & Brannon, 2013). These correlations also hold even when controlling for dozens of other possible variables, including intelligence, working memory, vocabulary, and speed of processing (Feigenson et al., 2013; Halberda et al., 2008) and predict math-specific learning deficits (Mazzocco, Feigenson, & Halberda, 2011; Piazza et al., 2010). Indeed, two separate meta-analyses have confirmed a reliable but small effect between the ANS and symbolic mathematics (Chen & Li, 2014; Schneider et al., 2017). Recently, several labs have shown that temporary manipulations of the ANS—whether through short-term or long-term training—have carryover effects to symbolic math scores, but not to other domains (Hyde, Khanum, & Spelke, 2014; Park & Brannon, 2013, 2014; Wang, Odic, Halberda, & Feigenson, 2016), though other studies have failed to find such carry-over effects (Lindskog & Winman, 2016; Merkley, Matejko, & Ansari, 2017; Szűcs & Myers, 2017).

What mechanisms might explain the observed relationship between ANS precision and symbolic mathematics? Classically, researchers have focused on the role that the ANS might play in helping children acquire early meanings for number words (e.g., Shusterman, Slusser, Halberda, & Odic, 2016), or on the role the ANS might play in helping children roughly estimate the range of possible answers on simple math problems (e.g., Feigenson et al., 2013). More recently, researchers have pointed to a potentially key link between mathematical reasoning and numerical metacognition: the ability to decide how *confident* or *certain* observers are in their number representations (Bellon, Fias, & De Smedt, 2019). This could also apply to the ANS because the ANS generates not only an estimate for which of two collections of dots is more numerous, but also generates a confidence signal corresponding to the uncertainty of this decision (Halberda & Odic, 2014; Vo, Li, Kornell, Pouget, & Cantlon, 2014). For example, after answering a simple number discrimination problem, children as young as five can appropriately bet on whether they answered correctly or incorrectly (Vo et al., 2014). Similarly, when given a choice of two number discrimination trials—one which children should be very certain on, and one that they should be less certain of—children as young as five can correctly select the trial they are more certain of getting correct (Baer, Gill, & Odic, 2018; Baer & Odic, 2019). Importantly, individual differences in children's ability to detect differences in certainty are at least partly independent of ANS precision itself and may constitute a kind of domain-general currency of perceptual confidence that can be used across many different domains (Baer et al., 2018; Baer & Odic, 2019; see also De Gardelle, Le Corre, & Mamassian, 2016; De Gardelle & Mamassian, 2014).

Research on symbolic mathematics has also examined the role of children's numerical metacognition in accounting for individual differences on various math tasks (e.g., Bellon et al., 2019; Rinne & Mazzocco, 2014; Vo et al., 2014). Symbolic number metacognition is often theorized to help children identify appropriate strategies for the task at hand, evaluate how much time they should invest in specific problems, and assess whether they are likely to have answered a problem correctly or incorrectly (Bellon et al., 2019; Rinne & Mazzocco, 2014). For example, Bellon et al. (2019) demonstrated that individual differences in 7-year-olds' numerical metacognition significantly and uniquely predicts the speed and accuracy of solving addition problems, even

beyond the contributions of IQ, executive function, and domain-general metacognitive abilities (e.g., knowledge of optimal memory strategies).

If individual differences in symbolic mathematics are uniquely predicted by numerical metacognition, and if the ANS generates an approximate numerical confidence signal, then one possible explanation for the correlation between these two abilities may be that individual differences in metacognitive sensitivity *mediate* the relationship between the ANS and symbolic math. Indeed, one previous finding suggests that individual differences in ANS metacognition correlate with a standardized symbolic math test (the Test of Early Mathematical Ability [TEMA-3]) which has also previously been shown to robustly correlate with ANS precision (Vo et al., 2014). However, this previous work did not concurrently measure individual differences in ANS precision, leaving open the possibility that ANS metacognition contributes to symbolic mathematics *independently* of the ANS itself, or that the observed correlation between the ANS and symbolic math may be entirely explained by the correlation between ANS precision and ANS metacognition.

The existing work relating symbolic and non-symbolic numerical metacognition to math to date has also relied on measures of metacognitive *calibration*: how well children's internal certainty states match their actual accuracy. However, it is well known that measures of calibration combine metacognitive *sensitivity* (how well children can discriminate between their own certainty states, also referred to as metacognitive *resolution* or *precision*) and metacognitive *bias* (a general tendency to over/under-estimate certainty), two distinct components of metacognitive ability (for discussion, see Nelson, 1984). One possibility not yet considered in the literature is that these components of metacognition may have different relationships with numerical and/or mathematical reasoning (see Winman, Juslin, Lindskog, Nilsson, & Kerimi, 2014). For instance, it may be more important for children to be well-attuned to their own certainty (sensitivity) than it is for children to be able to accurately report their certainty (bias), or vice versa.

One strategy to separate metacognitive sensitivity from metacognitive bias is to use statistics largely drawn from Signal Detection Theory (e.g., gamma, Nelson, 1984; A'ROC, Vo et al., 2014; d' , Maniscalco & Lau, 2012; Salles, Ais, Semelman, Sigman, & Calero, 2016), though these measures are not as effective if children do not report both high and low certainty with relatively equal frequency (a common occurrence given children's documented overconfidence in many tasks, including number reasoning, Destan & Roebbers, 2015; Nelson, 1984; Vo et al., 2014). Another strategy, which we adopt here, is to *experimentally* isolate metacognitive sensitivity by asking children to compare their certainty between two trials, rather than reporting their certainty in a single trial (see Nelson, 1984; Nelson & Narens, 1980). These "relative" metacognitive assessments have been used in memory and perception research for several decades, (Barthelmé & Mamassian, 2009; Butterfield, Nelson, & Peck, 1988; De Gardelle & Mamassian, 2014; Lipowski, Merriman, & Dunlosky, 2013), and have recently been shown to be appropriate for measuring children's ANS metacognitive sensitivity separately from their calibration abilities (Baer et al., 2018; Baer & Odic, 2019).

Here, we sought to replicate and extend the existing work on the relationship between ANS metacognition and symbolic mathematics by concurrently measuring individual differences in three abilities (Figure 1): 1) ANS precision, measured through a classic assessment asking children to report the more numerous of two arrays of dots (Halberda & Feigenson, 2008); 2) ANS metacognitive sensitivity, measured through a relative confidence task in which children decide which of two available trials they are more certain of getting right (Baer et al., 2018; Baer & Odic, 2019); and 3) the Test of Early Mathematics 3rd Edition (TEMA-3, Ginsburg & Baroody,

2003), an age-standardized measure of symbolic mathematics ability appropriate for preschool-aged children. Beyond seeking to replicate the previously documented relationship between ANS precision and the TEMA-3, we also tested whether this relationship might be mediated by individual differences in ANS metacognition sensitivity. Furthermore, because previous work has shown that ANS precision selectively predicts the *informal* scale of the TEMA-3 (e.g., children's counting and basic arithmetic skills), over the *formal* scale (e.g., more advanced abilities such as word problems; Libertus, Feigenson, & Halberda, 2013) our secondary goal was to test whether ANS metacognition contributes separately to formal vs. informal components of the TEMA-3.

Method

Participants

A total of 72 four- to six-year-olds participated in the study ($M = 5; 5$ [years;months], range = 4;0 – 6;11, 25 4-year-olds, 26 5-year-olds, 21 6-year-olds, 38 girls). Two additional children were excluded for not completing the TEMA-3. We aimed to test 60 participants (roughly 20 per age group); additional participants are the result of a change in procedure to include rewards for accuracy, which we decided to include in the final sample for additional power¹. All children were tested in a sound-attenuated room in the Centre for Cognitive Development at the University of British Columbia, Vancouver, Canada.

On average, parents reported that children in this sample heard English about 83% of the time ($SD = 19\%$, range = 30% – 100%, data available for 71 of 72 children); the most common second languages were French (23% of the sample), Mandarin (18%), and Cantonese (17%, all others < 10%). Children's ethnicities were predominantly Caucasian (24 of 58 reported), East and Southeast Asian (18) or Mixed (13, all others < 10%). Maternal education (a proxy for socio-economic status) was mostly at the University level (1 Primary school, 3 High School, 8 College/Trade School, 26 Undergraduate, 10 Master's, 3 Doctoral, and 8 Professional Degrees such as Law or Medicine, data unavailable for the remaining 13 children). We did not ask parents to report the grade level of their child, but we estimate based on their birthdays that 10 were in grade 1, 27 were in kindergarten, and the remaining 35 were in preschool or in-home care at the time of testing.

Materials and Procedures

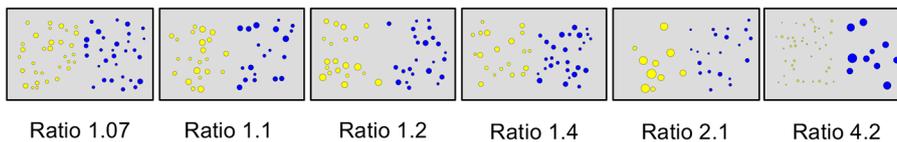
The study consisted of three tasks in a set order: 1) the ANS Discrimination Task; 2) the Relative ANS Metacognition Task; and 3) the TEMA-3. The ANS Discrimination Task and the Relative ANS Metacognition Task were both presented using an 11.3" Apple Air laptop computer using custom-made scripts in Psychtoolbox-3 (Brainard, 1997). These scripts are freely available online see the [Supplementary Materials](#) section. The TEMA-3 was subsequently administered by the researcher in the same room and followed the standardized testing guidelines set by the test. After completing the TEMA-3, children would "trade in" coins they earned during the study in exchange for a small prize (a book, stuffed animal, or shirt) and a diploma as a present for their participation (all children were given the same choice of prizes no matter how many coins they accrued).

ANS Discrimination Task

To measure the precision of each child's intuitive sense of number, we used a 30-trial dot comparison task used widely in the literature on the ANS (Halberda et al., 2008; Odic & Starr, 2018). Each trial consisted of a group of yellow dots on the left side of the screen and a group of blue dots on the right (see [Figure 1](#)) presented

for 1,200 ms, and children were subsequently asked to select the group that has more dots without counting. Dot area varied randomly in between 0.35° and 1.82° visual angle, measured at a viewing distance of 40 cm (though we did not restrict children's viewing distance, we estimate that most children sat about this distance from the monitor). The rectangle within which the dots were drawn subtended approximately $10.1^\circ \times 20.9^\circ$ visual angle. The cumulative area and density of the dots was controlled for across trials; on half the trials the cumulative area and density was Congruent with the side that had more dots (e.g., if the left side had twice as many dots, it also had twice the cumulative area), and on the other half the cumulative area and density was Incongruent with the side that had more dots (e.g., if the left side had twice as many dots, it also had half the cumulative area). To vary difficulty and draw out individual variability associated with the precision of children's ANS, we presented 5 numerical ratios: 1.13 (e.g., 8 yellow dots and 9 blue dots), 1.33, 1.5, 2.0, and 3.0. Children received pre-recorded accurate positive or negative feedback from the computer (e.g., "That's right!" or "Oh, that's not right!") to motivate children to complete as many trials as possible, and children were rewarded with a coin in their cup for each correct answer. Occasionally, the experimenter would give feedback to encourage the child to stay engaged in the task (e.g., "That's okay, let's do another one!"). To avoid the influence of motor development on the results, children either verbally identified or pointed to the set with more dots, while the experimenter pushed a corresponding button to indicate their answer. Spearman-Brown split-half corrected reliability for this task was .65.

a. Number Discrimination Trials



b. Relative ANS Metacognition Task

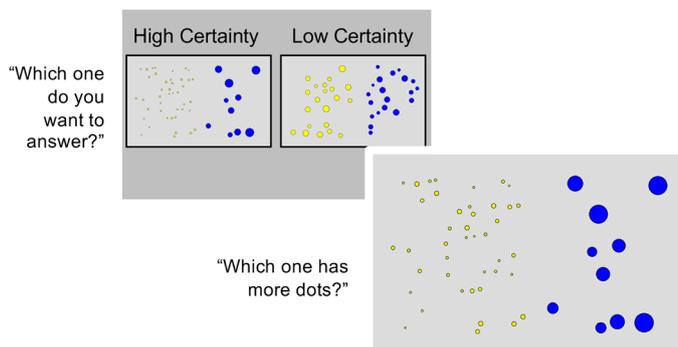


Figure 1. Sample stimuli used in the study.

Note. In the Number Discrimination Task (Section a), children indicated which color had more dots. In the Relative Approximate Number System Metacognition Task (Section b), children selected the question they expected to answer correctly, then answered only that question.

Relative ANS Metacognition Task

This task was modeled after existing work in adults (e.g., Barthelmé & Mamassian, 2009; De Gardelle et al., 2016; De Gardelle & Mamassian, 2014) and children (e.g., Baer et al., 2018; Butterfield et al., 1988; Lipowski et al., 2013), which measures individual differences in metacognitive sensitivity. To experimentally isolate metacognitive sensitivity from other components of metacognition, including calibration/bias, children are not

asked to provide a report of their confidence in a single item, but rather select one of two items that they feel more confident in answering (for discussion, see Nelson, 1984). This task logic stems from Signal Detection Theory, and is also the origin of the ANS Discrimination Task we use (we can isolate numerical reasoning from under-estimation biases or a lack of knowledge of number words by asking children to indicate which of two sets is more numerous). Previous work has shown that this paradigm reliably measures individual differences in certainty sensitivity (the smallest difference in certainty that children can identify) while alleviating issues with over- or under-confidence; since children are always selecting the trial they are *more* certain in, their general tendency to be over- or under-confident across the board does not impact performance on the task (see Baer et al., 2018).

In each of 30 trials, children were shown two screenshots of dot comparison trials (as in the ANS Discrimination Task described above, except that each screenshot subtended approximately 5.87° by 9.38° visual angle) that varied in numerical ratio, and were asked to select one question to answer (see Figure 1). Children would verbally identify or point toward the side containing the trial that they wished to attempt. Given previous work (Baer et al., 2018; Baer & Odic, 2019) and the coin reward structure, we expected children to be motivated to choose the higher certainty (i.e., easier) of the two dot comparison questions. Afterwards, the trial children identified would zoom in and cover the entire screen, and children would be asked to identify which set—the blue or the yellow—had more dots, in the identical manner to the ANS Discrimination Task above. Children received no feedback for their trial choice (i.e., the metacognitive part of the task), but did receive pre-recorded positive and negative feedback from the computer for the dot comparison trial they chose and would receive a coin if they answered the selected trial correctly. As in the ANS Discrimination Task, the experimenter would occasionally give feedback to encourage the child to stay engaged in the task (e.g., “That’s okay, let’s do another one!”).

To measure which children were better able to distinguish smaller and smaller differences in their certainty, we varied the difference between the two presented trials by controlling their “metaratio” (i.e., the ratio of the ratios for the two presented dot comparison trials). On each trial, children were presented with one of five metar ratios: 1.1 (e.g., ratio 1.33 vs. ratio 1.2), 1.25, 1.5, 2.0, and 4.0. While the large metar ratios, such as 4.0, present children with an extremely easy versus an extremely difficult trial, the smaller metar ratios, such as 1.1, present children with two trials that are both intermediately difficult and that could only be differentiated if children are representing certainty at a very narrow grain. Spearman-Brown split-half corrected reliability for this task was .55. The dependent variable in this task was accuracy in choosing the easier (larger) ratio.

TEMA-3

To assess children’s math skills, all children received Form A of the TEMA-3 (Ginsburg & Baroody, 2003), which has been commonly used to examine the relationship between the ANS and math performance (e.g., Halberda et al., 2008; Libertus et al., 2013; Vo et al., 2014). The TEMA-3 is a standardized test of mathematics ability normed for children age 3–8 years that is administered verbally by the experimenter with simple materials like tokens or flashcards. Following the standardized guidelines for the TEMA-3, the experimenter would begin the test at an age-normed question and proceed to ask harder and harder questions until the child hit a ceiling level of performance; subsequently, the experimenter would ask easier questions until the child hit the base level. Raw scores on the test were converted to an age-normalized score centered at 100. In addition, we calculated two subscale scores: Informal Score, which includes basic counting skills, number comparisons using spoken words or objects, and informal calculation using objects; and the Formal Score, which relies on concepts that

are typically only learned through schooling, such as reading and writing numbers, memorized facts about common calculations, and solving written equations. Because no age norms exist for these subscales, we report the average number of questions correctly solved by the child; all analyses involving Informal and Formal scores instead statistically control for age.

Results

ANS Discrimination Task

Consistent with past work on children's developing sense of number (Halberda & Feigenson, 2008; Odic, 2018), children chose the more numerous side 81% of the time in the ANS Discrimination Task ($SD = 11\%$), significantly above chance of 50%, one sample $t(71) = 24.91$, $p < .001$, $d = 2.94$. Children's choices were not impacted by the congruency between the cumulative area and density of the dots (Congruent: $M = 81\%$, $SD = 12\%$, Incongruent: $M = 81\%$, $SD = 13\%$, paired $t(71) = 0.47$, $p = .638$), suggesting that children did not reliably use the non-numeric cues in this task; all future analyses collapse across these two trial types. A Greenhouse-Geisser corrected repeated-measures analysis of covariance (ANCOVA) with Ratio as the independent variable (IV) and Age as a covariate showed that accuracy on this task was ratio-dependent, $F(3.31, 231.98) = 35.23$, $p < .001$, $\eta_p^2 = .34$ (see Figure 2a), and that children were more accurate with age, $F(1, 70) = 17.65$, $p < .001$, $\eta_p^2 = .20$. There was no interaction between age and ratio, $F(3.31, 231.98) = 1.59$, $p = .187$, $\eta_p^2 = .02$. These results replicate a large body of work on developing non-symbolic number representations, including its ratio-dependency and developmental trajectory (for reviews, see Odic & Starr, 2018; Szudlarek & Brannon, 2017).

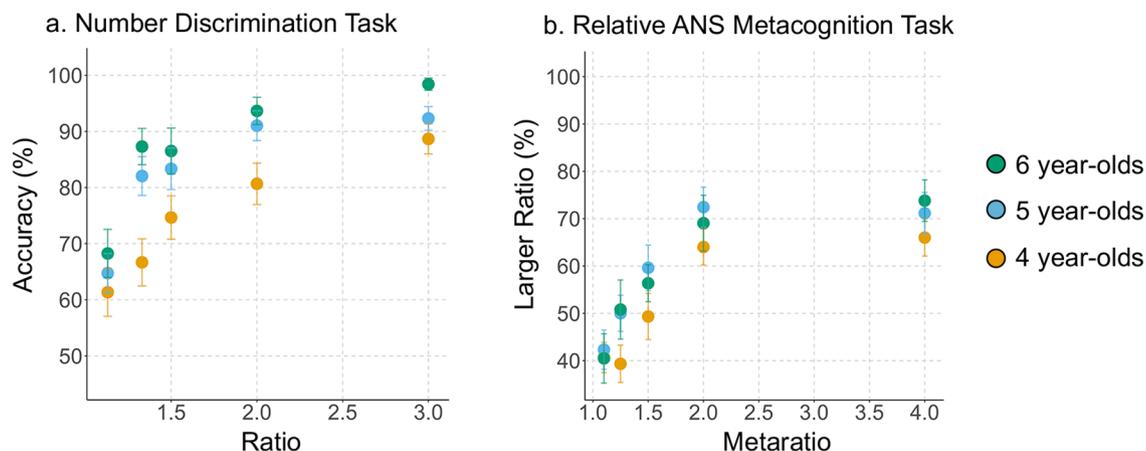


Figure 2. Accuracy at each ratio on the Number Discrimination Task, and at each metaratio on the Relative Approximate Number System Metacognition Task.

Note. Ages are binned for presentation, but treated continuously in our analyses. Error bars represent 1 SE.

Relative ANS Metacognition Task

In the Relative ANS Metacognition Task, children chose the easier question 56% of the time ($SD = 14\%$), a rate significantly higher than chance of 50%, $t(71) = 3.88$, $p < .001$, $d = 0.46$. This is consistent with past work using this paradigm and suggests that children chose the trial they were more confident in answering correctly.

Also consistent with past work, we find large individual differences in this task: while some children could only distinguish very large differences between the ratios (i.e., only between trials they were “very sure” vs. “not at all sure”), other children could reliably distinguish even small differences (i.e., trials they were “very sure” vs. “somewhat sure”, Baer et al., 2018; Baer & Odic, 2019)ⁱⁱ. A Greenhouse-Geisser corrected repeated-measures ANCOVA with Metaratio (i.e., the ratio between ratios) as the IV and Age as a covariate showed that accuracy on this task was ratio-dependent, $F(3.48, 243.70) = 33.21, p < .001, \eta_p^2 = .32$, see Figure 2b, but that accuracy did not improve with Age, $F(1, 70) = 2.86, p = .095, \eta_p^2 = .04$; there was also no significant Metaratio by Age interaction, $F(3.48, 243.70) = 0.80, p = .511, \eta_p^2 = .01$. We also found no correlation between children’s choice of trials on the ANS Metacognition Task and ANS accuracy when controlling for age, $r(69) = .06, p = .598$, consistent with previous reports that children’s metacognition on ANS tasks taps into more domain-general metacognitive components (Baer et al., 2018). These results broadly replicate previous work showing that as the difference in certainty between two trials grows, children are increasingly sensitive to this difference.

TEMA-3

Children’s TEMA-3 age-normed scores averaged 107.75 ($SD = 13.14$), reflecting a slightly above-norm sample, one-sample $t(71) = 5.01, p < .001, d = 0.59$. Both the Informal ($M = 23.38, SD = 9.27$) and Formal scores ($M = 5.86, SD = 5.67$) strongly correlated with age, Informal: $r(70) = .81, p < .001$, Formal: $r(70) = .66, p < .001$; note that the TEMA-3 does not provide age-norms for these two scales, and hence we only report the raw scores for these two subscales.

Regression Analyses

Next, we examined whether ANS accuracy and Relative ANS Metacognition account for variability in TEMA-3 performance, first for the age-normed TEMA score, and then for the Informal and Formal subscales. We replicate the previously documented relationship between ANS precision and children’s TEMA-3 scores (see Table 1): a hierarchical linear regression entering Age and ANS accuracy in Step 1 significantly predicted children’s standardized TEMA-3 scores, with ANS accuracy as the only significant predictor. However, we found that the addition of ANS Metacognitive accuracy in Step 2 did not add any further predictive power, and that the beta coefficient for this factor is not significantly different from zero (Table 1). In other words, when controlling for individual differences in Age and ANS, non-symbolic ANS Metacognition does not contribute additional explanatory power to the TEMA-3.

Previous work has shown that the relationship between the ANS and formal mathematics may be selective for only the Informal subscale of the TEMA-3 (Libertus et al., 2013). To confirm this, we conducted a mixed-level linear regression over the raw formal and informal TEMA scores with Age and ANS accuracy as fixed effects and subjects as random slopes; we dummy-coded the TEMA scores into a Type variable (i.e., “Formal” or “Informal”), and added it as an interaction term for ANS accuracy. Likelihood ratio tests were used for model significance testing: Step 1 was compared to an intercept-only model, Step 2 was compared to the model from Step 1. In Step 1 (ANS accuracy, Type, and Age), the fixed effects accounted for 83% of the variance, with no significant effect of Type or ANS, but a significant effect of Age and a significant ANS \times Type interaction (see Table 1). Estimated marginal means showed that this significant interaction was carried by a significantly positive slope for ANS accuracy in the Informal condition ($M = 0.28, SE = 0.06$) compared to the Formal condition ($M = -0.01, SE = 0.06$); this contrast was significant, $t(70) = 4.94, p < .001$. Therefore, we replicate work showing that ANS accuracy significantly predicts Informal but not Formal TEMA subscale scores, even

when controlling for age. Furthermore, adding Metacognitive accuracy in Step 2 did not add any additional explanatory power (see Table 1). Therefore, even when examining the two subscales separately, we find that Metacognitive accuracy does not contribute any explanatory power beyond the contributions of ANS accuracy and Age.

Table 1

Hierarchical Regressions Predicting TEMA-3 Performance

Model	Model Fit				Predictor	Coefficients		
	R^2	ΔR^2	F	p		β	t	p
DV: TEMA-3								
Step 1	.16	.16	6.41	.003	Age	0.001	0.01	.991
					ANS	0.40	3.16	.002
Step 2	.17	.01	0.95	.333	Age	-0.02	-0.14	.892
					ANS	0.39	3.13	.003
					Metacognition	0.11	0.98	.333
Model	R^2	ΔR^2	χ^2	p	Predictor	β	t	p
DV: Formal/Informal Score								
Step 1	.82	.82	844.65	.001	Age	0.43	9.04	.001
					ANS	-0.007	-0.15	.884
					Type (Informal)	-0.24	-1.18	.240
					ANS \times Type (Informal)	1.02	4.94	.001
Step 2	.82	.001	2.51	.29	Age	0.42	8.71	.001
					ANS	-0.003	-0.07	.943
					Type (Informal)	-0.36	-1.65	.104
					ANS \times Type (Informal)	0.97	4.72	.001
					Metacognition	-0.02	-0.32	.751
					Metacognition \times Type (Informal)	0.17	1.47	.145

Note. DV = Dependent Variable; ANS = Approximate Number System; TEMA-3 = Test of Early Mathematical Ability. All beta values are standardized.

Mediation Analyses

Although our initial analytic plan was to conduct a full mediation analysis between ANS accuracy, Metacognitive accuracy, and the TEMA-3, we did not do so as we found that Metacognitive accuracy was not predicted by the ANS, $r(69) = .06$, $p = .598$, nor did it predict TEMA-3 scores when considered alone, $r(69) = .13$, $p = .271$, or when additionally controlling for ANS accuracy, $\beta = 0.11$, $p = .333$. As a result, Metacognitive accuracy cannot be entered into a formal mediation analysis with these variables (Baron & Kenny, 1986).

Additional Analyses of the Relative ANS Metacognition Task

Given that the interpretation of our [null] result is contingent on the Metacognitive Task accurately measuring variability in ANS metacognitive sensitivity, we conducted additional analyses for this measure.

First, to check that children were paying attention during the Metacognitive Task and not responding randomly on every trial, we examined whether the ANS portion of the task (i.e., when children would answer their chosen question) followed similar patterns to the ANS Discrimination task. We found that children chose the side with more dots 83% of the time ($SD = 10\%$), well above chance of 50%, $t(71) = 28.74$, $p < .001$, $d = 3.92$. In fact, a linear regression with the accuracy on the chosen ANS trials from the Metacognitive Task and

Age predicted children's TEMA-3 scores, $R^2 = .10$, $F(2, 69) = 3.77$, $p = .03$; $\beta_{\text{ANS}} = 0.29$, $p = .027$; $\beta_{\text{Age}} = 0.06$, $p = .645$, showing that the ANS component of the Metacognitive Task captured meaningful variability in children's performance.

Next, because children in this sample were only slightly more likely than chance to select the larger metaratio, we also re-ran our regression analyses with the subset of participants whose performance was significantly above chance (see Inglis, Attridge, Batchelor, & Gilmore, 2011 for a similar approach). We removed all children who chose the larger metaratio 50% or less of the time, leaving a sample of 42 children. Nevertheless, we once again did not find that adding Metacognitive accuracy into a regression with ANS and Age significantly changed R^2 , including for formal vs. informal scores (all $\Delta R^2 < .05$, all p 's $> .143$). We also again failed to find that Metacognitive accuracy correlated with the ANS, $r(39) = -.09$, $p = .565$ or any of the TEMA-3 scores (all r 's $< .19$, all p 's $> .240$).

Finally, as can be seen in Figure 2b, children's performance on the Metacognitive Task varied with metaratio (just like their ANS performance varied by numerical ratio). Because the lowest metarations may have been too difficult for children and might have been introducing noise into the measure, we also decided to examine children's performance on the Metacognitive task by only taking the largest three metarations (1.5, 2.0, and 4.0) into consideration. When looking at these metarations together, children chose the larger (more certain) ratio on 65% of trials ($SD = 18\%$), above chance of 50%, one-sample $t(71) = 6.98$, $p < .001$, $d = 1.72$, showing that children's responses were based on their metacognitive states for these metarations. But, once again, this had no impact on the relationship between Metacognitive Task performance and ANS accuracy, $r(69) = .10$, $p = .402$, or any TEMA-3 measure, all r 's $> .14$, all p 's $> .250$. Similarly, there was no independent effect of Metacognitive accuracy in the hierarchical regression already including age and ANS as predictors for TEMA-3 scores, Informal scores, or Formal scores, all $\Delta R^2 < .01$, all p 's $> .365$.

Overall, then, even when looking at these subsamples of our data that maximize children's performance on the metacognitive task, we replicate the main findings that children chose trials to complete using a sense of certainty and answered those trials with their ANS; while the latter component was predictive of symbolic math ability, the former was not.

Discussion

What factors explain the observed relationship between individual differences in children's intuitive number sense and symbolic mathematics? Here, we tested whether non-symbolic/ANS numerical metacognition—that is, the sensitivity of children's ability to decide when they can easily discriminate between ANS trials—might mediate the reported relationship between ANS precision and performance on a standardized test of symbolic mathematics (TEMA-3). While we replicated the previously documented relationship between ANS precision and the TEMA-3, we failed to find any reliable relationship between ANS metacognitive sensitivity and the TEMA-3. We therefore conclude that individual differences in ANS metacognitive sensitivity may not contribute to the relationship between ANS precision and the TEMA-3.

If, as noted in the Introduction, the ANS generates a metacognitive signal (Baer & Odic, 2019; Halberda & Odic, 2014), and if symbolic math performance is at least partly predicted by more general numerical metacognition (Bellon et al., 2019; Rinne & Mazzocco, 2014; Vo et al., 2014), why did we fail to observe a relationship

between these variables? We see at least two possible explanations for this failure. One is that our relative metacognition task failed to measure children's metacognitive sensitivity, or at least failed to measure it with sufficient power to detect an existing relationship. This explanation, however, is inconsistent with the fact that children performed above chance at choosing the more certain trial, particularly on the largest metaratio where the difference in certainty was larger. If children were answering randomly, or succeeding on the task by computing a quantity other than confidence, we would not expect either of these to hold. Moreover, when we looked at only children who performed above chance, and when we removed the two lowest-performing metaratio, the relationship with TEMA-3 scores did not change.

An alternative explanation is that perceptual metacognition is not directly relevant for symbolic mathematics, even if symbolic number metacognition is. In previous work using this relative certainty measure, children's performance on a relative confidence ANS task strongly correlated with performance on a relative confidence surface area and emotion tasks, even though these three dimensions are perceptually independent of each other (Baer et al., 2018). In other words, and consistent with similar suggestions in the adult literature (De Gardelle et al., 2016; De Gardelle & Mamassian, 2014), perceptual confidence may be represented *domain-generally*, and may not be tied to specific representations that instantiate it. Individual differences in the relative ANS metacognition task may therefore broadly tap into general perceptual metacognition, rather than specific numerical metacognition. And, since previous work has demonstrated that number-specific metacognitive abilities are more predictive of symbolic math performance than domain-general metacognitive ones (Bellon et al., 2019), one potential explanation for our finding may be that the ANS Metacognition Task only taps into the more general differences in perceptual metacognition that are not as relevant for early math abilities. Furthermore, it is also possible that while symbolic number metacognition is relevant for mathematics, non-symbolic number metacognition may not be, whether tapped into by our task or not.

One challenge for this explanation is that a positive correlation between ANS metacognition and the TEMA-3 has been previously reported by Vo et al. (2014). One key difference between their work and ours is the nature of the metacognitive task. In Vo and colleagues' work, children's metacognition was assessed through an "absolute" confidence task: children were asked to rate their confidence as either high or low by betting coins depending on whether they believed they answered the dot comparison trial correctly or incorrectly. Beyond potentially tapping into individual differences in impulse control, risk aversion, and other executive function differences, absolute confidence tasks cannot always separate metacognitive sensitivity from metacognitive bias: a general tendency children have to constantly indicate high or low confidence (Baer & Odic, 2019; Butterfield et al., 1988; Lipowski et al., 2013; Nelson, 1984). Therefore, one possible explanation that accounts for all the available evidence is that children's metacognitive sensitivity alone (what we experimentally isolate in our measure) is not related to their TEMA-3 performance, but their metacognitive *calibration* is related (as found in past work, Bellon et al., 2019; Rinne & Mazzocco, 2014; Vo et al., 2014; and see Winman et al., 2014 for further discussion of components of metacognition). Future work can test this possibility by simultaneously measuring children's metacognitive sensitivity and metacognitive calibration, ideally through a mixture of both relative and absolute certainty tasks.

Although we did not find a significant correlation between ANS metacognitive sensitivity and symbolic mathematics, we did successfully replicate the correlation between ANS precision and the TEMA-3 (Halberda et al., 2008; Starr et al., 2013; for review, see Szkludlarek & Brannon, 2017), including the stronger relationship between the ANS and the *informal* questions on the TEMA-3 compared to formal ones (Libertus et al., 2013).

We found that this relationship held even when controlling for age and individual differences in ANS metacognitive sensitivity. While there has been a significant amount of debate about the validity of the correlation between the ANS and mathematical abilities (e.g., Gilmore et al., 2013; Szűcs, Nobes, Devine, Gabriel, & Gebuis, 2013), two recent meta-analyses have shown that there is a small, but reliable effect between the ANS and mathematical abilities (Chen & Li, 2014; Schneider et al., 2017). Our data contributes to this literature, and generally aligns with the size of the effect that has been documented by these meta-analyses.

In conclusion, the primary aim of this work was to test whether ANS metacognitive sensitivity correlates with formal mathematics, and if so, whether it acts as a mediating variable between ANS precision and the TEMA-3. We failed to find any significant relationship between ANS metacognitive sensitivity and the TEMA-3, suggesting either that metacognitive calibration may play a larger role or that metacognitive differences in the ANS do not robustly contribute to number-specific metacognitive abilities that are thought to be critical for symbolic math performance.

Notes

- i) Thirteen children in the sample were not rewarded with coins because this change was added after testing began to try and boost children's motivation to complete the task. Results do not change in direction or magnitude with these children removed.
- ii) Ten children in our sample consistently chose the smaller ratio, a similar rate to other reported results using this paradigm (Baer et al., 2018; Baer & Odic, 2019). Results reported here do not differ in direction or magnitude if these children are removed from the sample.

Funding

This work was supported by an Insight Grant from the Social Sciences and Humanities Research Council of Canada (SSHRC) to DO, and a SSHRC Joseph-Armand Bombardier Canada Graduate Scholarship to CB.

Competing Interests

The authors have declared that no competing interests exist.

Acknowledgments

Thanks to Sally Poon, Harris Wong, Chantane Yeung, Brionne McWilliams, and to the children and families who participated.

Supplementary Materials

The raw data and custom-made Psychtoolbox-3 program scripts for the ANS Discrimination Task and the Relative ANS Metacognition Task are freely available online at (for access see Index of [Supplementary Materials](#) below).

Index of Supplementary Materials

Baer, C., & Odic, D. (2020). *Supplementary materials to "The relationship between children's approximate number certainty and symbolic mathematics"* [Research data and code]. OSF. <http://www.doi.org/10.17605/osf.io/82phe>

References

- Baer, C., Gill, I. K., & Odic, D. (2018). A domain-general sense of confidence in children. *Open Mind: Discoveries in Cognitive Science*, 2, 86-96. https://doi.org/10.1162/opmi_a_00020
- Baer, C., & Odic, D. (2019). Certainty in numerical judgments develops independently of the Approximate Number System. *Cognitive Development*, 52, Article 100817. <https://doi.org/10.1016/j.cogdev.2019.100817>
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182. <https://doi.org/10.1037/0022-3514.51.6.1173>
- Barthelmé, S., & Mamassian, P. (2009). Evaluation of objective uncertainty in the visual system. *PLoS Computational Biology*, 5, Article e1000504. <https://doi.org/10.1371/journal.pcbi.1000504>
- Bellon, E., Fias, W., & De Smedt, B. (2019). More than number sense: The additional role of executive functions and metacognition in arithmetic. *Journal of Experimental Child Psychology*, 182, 38-60. <https://doi.org/10.1016/j.jecp.2019.01.012>
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433-436. <https://doi.org/10.1163/156856897X00357>
- Butterfield, E. C., Nelson, T. O., & Peck, V. (1988). Developmental aspects of the feeling of knowing. *Developmental Psychology*, 24, 654-663. <https://doi.org/10.1037/0012-1649.24.5.654>
- Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica*, 148, 163-172. <https://doi.org/10.1016/j.actpsy.2014.01.016>
- De Gardelle, V., Le Corre, F., & Mamassian, P. (2016). Confidence as a common currency between vision and audition. *PLoS One*, 11, Article e0147901. <https://doi.org/10.1371/journal.pone.0147901>
- De Gardelle, V., & Mamassian, P. (2014). Does confidence use a common currency across two visual tasks? *Psychological Science*, 25, 1286-1288. <https://doi.org/10.1177/0956797614528956>
- Dehaene, S. (2009). Origins of mathematical intuitions. *Annals of the New York Academy of Sciences*, 1156, 232-259. <https://doi.org/10.1111/j.1749-6632.2009.04469.x>
- Destan, N., & Roebbers, C. M. (2015). What are the metacognitive costs of young children's overconfidence? *Metacognition and Learning*, 10, 347-374. <https://doi.org/10.1007/s11409-014-9133-z>
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8, 307-314. <https://doi.org/10.1016/j.tics.2004.05.002>
- Feigenson, L., Libertus, M. E., & Halberda, J. (2013). Links between the intuitive sense of number and formal mathematics ability. *Child Development Perspectives*, 7, 74-79. <https://doi.org/10.1111/cdep.12019>
- 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, Article e67374. <https://doi.org/10.1371/journal.pone.0067374>

- Gilmore, C., McCarthy, S. E., & Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature*, *447*, 589-591. <https://doi.org/10.1038/nature05850>
- Ginsburg, H. P., & Baroody, A. J. (2003). *Test of early mathematics ability—Third edition*. Austin, TX, USA: PRO-ED.
- 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*, 1457-1465. <https://doi.org/10.1037/a0012682>
- Halberda, J., Mazocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*, 665-668. <https://doi.org/10.1038/nature07246>
- Halberda, J., & Odic, D. (2014). The precision and internal confidence of our approximate number thoughts. *Evolutionary Origins and Early Development of Number Processing*, *1*, 305-333. <https://doi.org/10.1016/B978-0-12-420133-0.00012-0>
- Hyde, D. C., Khanum, S., & Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition*, *131*, 92-107. <https://doi.org/10.1016/j.cognition.2013.12.007>
- Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: But only in children. *Psychonomic Bulletin & Review*, *18*, 1222-1229. <https://doi.org/10.3758/s13423-011-0154-1>
- 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>
- Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Numerical approximation abilities correlate with and predict informal but not formal mathematics abilities. *Journal of Experimental Child Psychology*, *116*, 829-838. <https://doi.org/10.1016/j.jecp.2013.08.003>
- Libertus, M. E., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica*, *141*, 373-379. <https://doi.org/10.1016/j.actpsy.2012.09.009>
- Lindskog, M., & Winman, A. (2016). No evidence of learning in non-symbolic numerical tasks—A comment on Park and Brannon (2014). *Cognition*, *150*, 243-247. <https://doi.org/10.1016/j.cognition.2016.01.005>
- Lipowski, S. L., Merriman, W. E., & Dunlosky, J. (2013). Preschoolers can make highly accurate judgments of learning. *Developmental Psychology*, *49*, 1505-1516. <https://doi.org/10.1037/a0030614>
- Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*, 422-430. <https://doi.org/10.1016/j.concog.2011.09.021>
- Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development*, *82*, 1224-1237.
- McCrink, K., & Spelke, E. S. (2016). Non-symbolic division in childhood. *Journal of Experimental Child Psychology*, *142*, 66-82. <https://doi.org/10.1016/j.jecp.2015.09.015>
- Merkley, R., Matejko, A. A., & Ansari, D. (2017). Strong causal claims require strong evidence: A commentary on Wang and colleagues. *Journal of Experimental Child Psychology*, *153*, 163-167. <https://doi.org/10.1016/j.jecp.2016.07.008>

- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, 95, 109-133. <https://doi.org/10.1037/0033-2909.95.1.109>
- Nelson, T. O., & Narens, L. (1980). A new technique for investigating the feeling of knowing. *Acta Psychologica*, 46, 69-80. [https://doi.org/10.1016/0001-6918\(80\)90060-8](https://doi.org/10.1016/0001-6918(80)90060-8)
- Odic, D. (2018). Children's intuitive sense of number develops independently of their perception of area, density, length, and time. *Developmental Science*, 21, Article e12533. <https://doi.org/10.1111/desc.12533>
- Odic, D., & Starr, A. (2018). An introduction to the approximate number system. *Child Development Perspectives*, 12, 223-229. <https://doi.org/10.1111/cdep.12288>
- Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychological Science*, 24, 2013-2019. <https://doi.org/10.1177/0956797613482944>
- Park, J., & Brannon, E. M. (2014). Improving arithmetic performance with number sense training: An investigation of underlying mechanism. *Cognition*, 133, 188-200. <https://doi.org/10.1016/j.cognition.2014.06.011>
- Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., . . . Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, 116, 33-41. <https://doi.org/10.1016/j.cognition.2010.03.012>
- Rinne, L. F., & Mazzocco, M. M. M. (2014). Knowing right from wrong in mental arithmetic judgments: Calibration of confidence predicts the development of accuracy. *PLoS One*, 9(7), Article e98663. <https://doi.org/10.1371/journal.pone.0098663>
- Salles, A., Ais, J., Semelman, M., Sigman, M., & Calero, C. I. (2016). The metacognitive abilities of children and adults. *Cognitive Development*, 40, 101-110. <https://doi.org/10.1016/j.cogdev.2016.08.009>
- Sasanguie, D., Defever, E., Maertens, B., & Reynvoet, B. (2014). The approximate number system is not predictive for symbolic number processing in kindergarteners. *Quarterly Journal of Experimental Psychology*, 67, 271-280. <https://doi.org/10.1080/17470218.2013.803581>
- Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & DeSmedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science*, 20, Article e12372. <https://doi.org/10.1111/desc.12372>
- Shusterman, A., Slusser, E., Halberda, J., & Odic, D. (2016). Acquisition of the cardinal principle coincides with improvement in approximate number system acuity in preschoolers. *PLoS One*, 11, Article e0153072. <https://doi.org/10.1371/journal.pone.0153072>
- Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences*, 110, 18116-18120.
- Szkudlarek, E., & Brannon, E. M. (2017). Does the approximate number system serve as a foundation for symbolic mathematics? *Language Learning and Development*, 13, 1-20. <https://doi.org/10.1080/15475441.2016.1263573>
- Szkudlarek, E., & Brannon, E. (2018). Non-symbolic division ability mediates the relation between visual number discrimination acuity and symbolic math skill. *Journal of Vision*, 18, 273-273. <https://doi.org/10.1167/18.10.273>

- Szűcs, D., & Myers, T. (2017). A critical analysis of design, facts, bias and inference in the approximate number system training literature: A systematic review. *Trends in Neuroscience and Education*, 6, 187-203. <https://doi.org/10.1016/j.tine.2016.11.002>
- Szűcs, D., Nobes, A., Devine, A., Gabriel, F. C., & Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. *Frontiers in Psychology*, 4, Article 444. <https://doi.org/10.3389/fpsyg.2013.00444>
- Vallortigara, G. (2017). An animal's sense of number. In J. Adams, P. Barmby, & A. Mesouid (Eds.), *The nature and development of mathematics: Cross disciplinary perspectives on cognition, learning and culture* (pp. 43–66). London, United Kingdom: Taylor & Francis.
- Vo, V. A., Li, R., Kornell, N., Pouget, A., & Cantlon, J. F. (2014). Young children bet on their numerical skills metacognition in the numerical domain. *Psychological Science*, 25, 1712-1721. <https://doi.org/10.1177/0956797614538458>
- Wang, J. J., Odic, D., Halberda, J., & Feigenson, L. (2016). Changing the precision of preschoolers' approximate number system representations changes their symbolic math performance. *Journal of Experimental Child Psychology*, 147, 82-99. <https://doi.org/10.1016/j.jecp.2016.03.002>
- Winman, A., Juslin, P., Lindskog, M., Nilsson, H., & Kerimi, N. (2014). The role of ANS acuity and numeracy for the calibration and the coherence of subjective probability judgments. *Frontiers in Psychology*, 5, Article 851. <https://doi.org/10.3389/fpsyg.2014.00851>