Developmental Review 72 (2024) 101131

Contents lists available at ScienceDirect

Developmental Review

journal homepage: www.elsevier.com/locate/dr

Review

Measuring the acuity of the approximate number system in young children

Attila Krajcsi [a,*], Dana Chesney [b], Krzysztof Cipora [c], Ilse Coolen [d], Camilla Gilmore [c], Matthew Inglis [c], Melissa Libertus [e], Hans-Christoph Nuerk [f,g,h], Victoria Simms [i], Bert Reynvoet [j,k]

[a] Department of Cognitive Psychology, Institute of Psychology, ELTE Eötvös Loránd University, Hungary
[b] Department of Psychology, St. John's University, Jamaica, NY, USA
[c] Centre for Mathematical Cognition, Loughborough University, Loughborough, UK
[d] Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Nijmegen, the Netherlands
[e] Department of Psychology, Learning Research and Development Center, University of Pittsburgh, USA
[f] Department of Psychology, University of Tuebingen, Tuebingen, Germany
[g] LEAD Graduate School and Research Network, University of Tuebingen, Tuebingen, Germany
[h] German Center for Mental Health (DZPG), partner site Tuebingen, Germany
[i] School of Psychology, Ulster University, UK
[j] Brain and Cognition, KU Leuven, Belgium
[k] Faculty of Psychology and Educational Sciences, KU Leuven Kulak, Belgium

ARTICLE INFO

Keywords:
Approximate Number System
Numerical development
Weber fraction
Reliability

ABSTRACT

The approximate number system (ANS) is a hypothesized mechanism responsible for the representation and processing of numerical information in an imprecise fashion. According to the predominant theory, the ANS is essential in solving simple numerical tasks such as comparing which of two quantities is numerically larger, and some research has indicated that individual differences in its acuity influence higher-level mathematical performance. Because of this far-reaching role of the ANS, it is essential to assess its acuity with measures that are reliable, and valid. The present work reviews and synthesizes many of the methodological problems that are relevant for measuring ANS acuity in young children. We discuss issues related to task comprehension, the role of non-numerical perceptual properties of the stimuli, the role of inhibition, and the appropriateness and reliability of the ANS acuity indices. Recommendations and open questions are summarized.

Introduction

The approximate number system (ANS) is thought to be responsible for the representation and processing of numerical information in an imprecise way. This system follows Weber's law, allowing ratio-dependent discrimination (Dehaene, 2007; Moyer & Landauer, 1967). It is evolutionarily old, seen in vertebrates as diverse as humans, chickens, and zebrafish (Brannon & Merritt, 2011; Parrish & Beran, 2022). The ANS is proposed to be the cornerstone of number processing (Dehaene, 1992; Moyer & Landauer, 1967; Nieder & Dehaene, 2009). It is implicated in many aspects of numerical cognition development (Feigenson et al., 2004; Odic & Starr, 2018;

* Corresponding author.
E-mail address: krajcsi@gmail.com (A. Krajcsi).

https://doi.org/10.1016/j.dr.2024.101131
Received 3 November 2023; Received in revised form 5 March 2024;
Available online 16 March 2024

Piazza, 2010). For example, it is proposed that individual differences in the acuity (or sensitivity) of the ANS, that is, how precisely numerical information is represented (Halberda et al., 2008), are related to the pace of learning the first symbolic numbers (Piazza, 2010) and also to more advanced mathematical abilities (Libertus et al., 2011, 2013a; Park & Brannon, 2013, 2014; Szkudlarek & Brannon, 2017).

However, it has long been debated whether and how the ANS relates to the development of mathematical abilities (Carey, 2009; Carey & Barner, 2019; Feigenson et al., 2004; Krajcsi et al., 2023). To resolve these debates and to empirically test the role of the ANS in mathematics development, a key issue is how ANS acuity should be measured (Dietrich et al., 2015; Leibovich et al., 2017; Mix et al., 1997; Szkudlarek & Brannon, 2017). Measuring ANS acuity may also be beneficial for the early detection of number processing problems and screening for mathematics learning difficulties (Butterworth, 2010).

Typically, ANS acuity is measured via a number comparison task. According to the ANS model, values of non-symbolic (such as arrays of dots) and symbolic numbers (such as Arabic numbers or number words) are processed by the ANS (Dehaene, 1992; Moyer & Landauer, 1967); therefore, both symbolic and non-symbolic number comparisons are used in the literature (Schneider et al., 2017). However, it is debated if symbolic and non-symbolic numbers are processed by the same system (Leibovich et al., 2017; Lyons et al., 2015), or if the ANS is fundamental at all in simple symbolic number processing (Krajcsi et al., 2022). Therefore, in the present paper, we focus on non-symbolic number comparison tasks as a tool to measure the ANS acuity. Although this type of magnitude comparison task is the most commonly used method to assess the ANS, there are alternative paradigms that have been employed, such as using habituation or preferential looking paradigms (i.e., methodologies based on looking time; e.g., Libertus & Brannon, 2010; Xu et al., 2005) or manipulating non-symbolic magnitudes in arithmetic operations (e.g., Chesney et al., 2021; Gilmore et al., 2011, 2014; Park & Brannon, 2013). However, tasks such as non-symbolic arithmetic tasks and magnitude comparison tasks show low correlations (in adults –Gilmore et al., 2011; in children – Coolen et al., 2022; but see Gilmore et al., 2014), indicating a potentially diverse range of divergent cognitive skills underlying the various tasks (Xenidou-Dervou et al., 2013). Thus, these tasks should not be used interchangeably and researchers should carefully consider the rationale for selecting a paradigm that deviates from the non-symbolic number comparison task. In addition, we focus on simultaneous non-symbolic comparisons, when both dot collections are presented at the same time (rather than sequentially) and they are spatially separated. There are other variants of non-symbolic comparison, and while the observations and recommendations of this paper may be valid for those variants, too, there are studies showing that other setups are not necessarily equivalent from the present viewpoint (see Braham & Libertus, 2018; Mou et al., 2023; Price et al., 2012, for differences and similarities between different variants of non-symbolic comparison tasks).

In non-symbolic comparison tasks, participants have to choose the more numerous of two sets of items that are presented too quickly to be counted precisely (Dehaene, 2007). Typically, the non-numerical perceptual properties of the stimuli (e.g., dot size, density, total area, convex hull) are partially controlled (e.g., Gebuis & Reynvoet, 2011), and/or the perceptual properties are accounted for when evaluating performance (e.g., DeWind et al., 2015). According to the psychophysical model (and in line with Weber's law), performance in the task depends upon the ratio of the to-be-compared values: Error rates are higher, and reaction times are longer for number pairs with a ratio closer to 1 (i.e., the well-established ratio effect) (Dehaene, 2007). Based on the comparison performance, the acuity of the ANS is often calculated as a Weber fraction, but alternative indices (such as mean accuracy or ratio effect slope – see below) are not uncommon.

Nevertheless, several aspects of ANS measurement methods have been questioned, such as how strongly non-numerical perceptual properties of the stimuli may influence non-symbolic task performance (DeWind et al., 2015; Gebuis & Reynvoet, 2011), whether domain-general capacities such as inhibition may influence those measurements (Fuhs & McNeil, 2013; Gilmore et al., 2013), whether different indices are reliable (DeWind et al., 2015; Lindskog et al., 2013; Maloney et al., 2010), and whether some indices are valid measurements of the Weber fraction (Chesney, 2018; Krajcsi, 2020), to name but a few. These methodological issues can fundamentally influence how empirical data should be interpreted and how appropriate diagnostic information can be (DeWind & Brannon, 2016; Leibovich et al., 2017; Negen & Sarnecka, 2015).

By and large, similar methods are used with both adults and children as young as three years of age. However, concerns about the validity of ANS assessment methods may be particularly important with young children. In this review, we define young children as those who are toddlers and pre-literate children (aged around 2–6 years old in many cultures). Some of the general issues may be especially pronounced with this age group and age-group-specific issues may also play a role.

The aim of this paper is to summarize and synthesize the methodological issues concerning the validity and reliability of ANS acuity measurement in young children and, when available, to provide recommendations and reasonable solutions for these issues. While methodological summaries for ANS measurement in general are available (Dietrich et al., 2015; Szkudlarek & Brannon, 2017), the present work focuses on the methodological issues that are specific to developmental studies focusing on early childhood. For a more general summary of the issues facing researchers conducting studies with older children and adults, the review of Dietrich and colleagues (2015) is an excellent starting point, while Szkudlarek and Brannon (2017) discuss other relevant details of the studies that measure the relation of the ANS and more advanced mathematical abilities. Note that while it is debated whether the ANS is an appropriate model for several phenomena (Krajcsi et al., 2022), the present work investigates how ANS acuity could be measured validly and reliably with a non-symbolic number comparison task. The present review focuses specifically on four key issues: (1) whether the instructions of the comparison task are understood by young children, (2) how non-numerical features and inhibition can influence the measurements, (3) which index should be used to operationalize ANS acuity, and (4) the reliability of ANS indices.

Make sure children understand the instructions

Many studies have demonstrated that infants are sensitive to numerical information (Feigenson et al., 2004; Odic & Starr, 2018).

For example, when presented with two streams of rapidly changing sets of dots, infants prefer to look to the stream where the number of dots changes compared to the stream where the images change but the quantity remains constant (Libertus & Brannon, 2010). Based on these studies, it is reasonable to assume that young children can rely on this ability in order to compare non-symbolic numerical stimuli.

However, this discrimination ability in infancy does not guarantee that young children will understand that they should use this ability to solve a similar comparison task. For studies with toddlers and children, a typical design is to ask the participants to identify which of two sets of items contains “more”, which goes beyond the automatic detection of a numerical feature during passive viewing. Recent work (Lindskog & Simms, 2021; Negen & Sarnecka, 2015) has raised the possibility that some children do not understand the instructions to “choose more/fewer dots” that are typical in non-symbolic number comparison tasks (although, see counter-examples, where even the youngest children seem to understand the task, e.g., in Halberda & Feigenson, 2008). Evidence for this comes from the finding that the performance of most two- and some three-year-old children was random, even for large-ratio pairs (e.g., 6 vs. 24 dots). (Note that such random performance can be detected when mean accuracy is around the chance level or when the sigmoid fit is inadequate – see more details in the Operationalizing section below.) Given that even infants can discriminate these large ratios, toddlers’ failure on such tasks is most logically attributable to a lack of comprehension of the task instructions rather than imprecision of their ANS.

Other works similarly highlight the difficulties that young children may face with understanding task instructions. For example, when children are asked to choose a picture containing the same number of items as appears in a sample picture, subset-knowers (i.e., children who can only meaningfully use the first few number words from their counting list when asked to give a specific number of objects, Wynn, 1990, 1992) are less likely to choose the pictures that match in terms of quantity than the pictures that match in terms of non-numerical perceptual properties (Slusser & Sarnecka, 2011). Indeed, issues of task demand have long been known to complicate the assessment of numerical competence. For example, Gelman (1972) showed that children who would typically fail classic Piagetian number conservation tasks can nevertheless demonstrate an understanding of number invariance, acting surprised at surreptitious quantity changes that seem to have occurred by “magic”.

A number of approaches to overcoming young children’s difficulties with task comprehension have been used in the past. One possible way to overcome this limited understanding of the task is to start with extremely large numerical ratios, which should be easy for the children to discriminate (i.e., ratio differences readily detectable by infants), and then to make the trials gradually harder (Negen & Sarnecka, 2015). In such a gradual training task (unlike in the classic variant), children demonstrated the expected ratio effect (Negen & Sarnecka, 2015). Another potential solution is to avoid using verbal instructions and, instead, to have the experimenter demonstrate easy (i.e., large-ratio) trials before presenting more difficult trials to the children (Lindskog & Simms, 2021). While the results of the latter method show that three-year-old children perform above the 50 % chance level, the ratio effect was not observed in their performance (Lindskog & Simms, 2021). The lack of the ratio effect questions whether this version of the magnitude comparison task actually measures ANS performance. However, because Lindskog and Simms’s test included only relatively large ratios (1:4, 1:2, and 2:3), it is possible that these ratios were too easy and resulted in a ceiling effect, thus, making it impossible to observe a ratio effect. Accordingly, further examination is needed before deciding whether nonverbal instructions are a viable alternative to measure numerical comparison performance or not. In other reports, the comparison task performance of even three-year-old children is above the chance level (Odic et al., 2013) when some training trials were used (though the nature of the training trials was unspecified in the report). Another possibility might be to employ behavior-based methodologies that do not require verbal instruction, such as surprise reaction to numerosity changes (e.g., Gelman, 1972), or looking time methodologies that have successfully been used to establish that preverbal infants and toddlers can detect numerosity differences (e.g., Navarro et al., 2018; Starkey et al., 1990; Xu et al., 2005). However, note that these alternative paradigms, together with their statistical evaluation methods, may provide ANS acuity in a different way than the two alternatives forced choice comparison task does; Therefore, the ANS acuities provided by different methods may not be compatible (see the specific differences of the methods and the limitations of using different methods together in Krajcsi et al., 2023). Overall, even if young children are able to understand the instructions of the comparison task, additional studies are still needed to identify which methods ensure that children do understand the task and do not perform randomly. Moreover, work is needed to identify whether and which specific subgroups of children (e.g., those with specific language impairment) may benefit the most from alternative instructions.

Some of these approaches propose that trials should be increasingly difficult in the non-symbolic number comparison task. A potential issue with these approaches is that the order of the trials has been observed to influence performance on ANS tasks (Odic et al., 2014). Specifically, ANS acuity is better when children start with easy trials and proceed to difficult trials, compared to the reverse condition. Therefore, the order of the trials (and paradigms that use trials increasing in difficulty) may influence the comparison performance. In other words, when trials get increasingly difficult to make the task more understandable, this modification may change the performance. Importantly, this order effect in difficulty was only observed when feedback was given after each trial; without feedback, the order effect disappeared (Odic et al., 2014). (See additional considerations of the role of feedback reported by Dietrich et al., 2015.).

To summarize, young children may not understand the “choose more/fewer dots” instruction in a number comparison task (Lindskog & Simms, 2021; Negen & Sarnecka, 2015), and consequently several previous studies may have failed to validly measure children’s numerical discrimination ability. Gradual training (assuming that no feedback is given after each trial, Odic et al., 2013), nonverbal instruction methods, or other to-be-identified methods could be appropriate means to measure numerical comparison performance in such children.

Consider non-numerical perceptual properties and inhibition

Many perceptual properties of sets of objects co-vary with number. For example, a more numerous set of dots usually has a larger cumulative surface area than a less numerous set of dots (this must be the case whenever the average dot size is identical between the sets). Similarly, when a given set of dots is evenly distributed within a fixed area (i.e., half of a computer screen), a more numerous set of dots will tend to have dots that are closer to each other than a less numerous set of dots. Thus, it is possible that non-numerical perceptual properties influence numerical comparison performance. Indeed, a crucial observation is the presence of congruency effects in non-symbolic number comparison: In a non-symbolic number comparison task, children's and adults' judgments are biased by the congruency or incongruency between numerical information and non-numerical perceptual properties. For instance, Gebuis and Reynvoet (2012) showed that adult participants are more accurate whenthe more numerous dot array has a larger convex hull, or a smaller aggregate surface. Similar observations have been made with young children (e.g., Defever et al., 2013).

It has been debated for a long time how these non-numerical perceptual properties in non-symbolic stimuli should be handled. Various methods have been developed to minimize the effects of perceptual properties (e.g., Gebuis & Reynvoet, 2011). However, such methods have their own limitations because the non-numerical perceptual features of the stimuli cannot be manipulated entirely independently of the numerical properties (DeWind et al., 2015). Another approach is to model the contribution of numerical and non-numerical properties on performance so that even if the effects of non-numerical properties are unavoidable, the specific effect of the numerical property on performance can be captured (DeWind et al., 2015; DeWind & Brannon, 2016). Two recent studies in young children adopting this latter strategy (Starr et al., 2017; Tomlinson et al., 2020) demonstrated that, although performance in young children is indeed affected by non-numerical properties, the numerical information remains the dominant feature driving performance, in line with the primary status of the numerical dimension for the task. However, this conclusion was contested by Aulet and Lourenco (2021). In a re-analysis of Tomlinson et al.'s (2020) data, that modeled additive cumulative area area (which is a better proxy for the perceived cumulative area), instead of mathematical cumulative area, Aulet and Lourenco (2021) observed that number and perceived area equally contributed to performance. In addition, and, more strikingly, area biased the number decisions more than the other way around, indicating that perceived area might be a more important feature than number. This echoes the findings of Hurewitz et al. (2006), who found Stroop-like interference of number on area judgments and vice versa, with area more strongly interfering with number judgments than number interfered with area judgments. In another study investigating the source of the improvement of comparison performance, it was found that, in three- to six-year-old children, performance improvement is mainly due to the fact that children start to focus on the numerical, instead of non-numerical, features of the stimuli, while the improvement of number detection sensitivity was less dominant (Piazza et al., 2018). Although more research on the interplay between numerical and non-numerical properties is needed, these recent studies make clear that the contribution of non-numerical properties needs to be extracted to arrive at a proper estimate of ANS acuity. This caution is even more important with young children, since they may rely more heavily on non-numerical properties (Piazza et al., 2018; Starr et al., 2017).

Because both relevant numerical and irrelevant non-numerical properties are processed, comparison performance may also depend on the ability to inhibit irrelevant information more generally (Clayton & Gilmore, 2015; Leibovich & Ansari, 2016). This inhibition is especially important in developmental studies when the relation between the ANS acuity and mathematical performance is investigated. It has been demonstrated that the observed ANS index and general mathematical ability may correlate because performance on non-symbolic number comparison tasks relies on inhibition, and inhibition correlates with mathematical abilities (Fuhs & McNeil, 2013; Gilmore et al., 2013). However, this observation cannot be consistently replicated. For example, Keller and Libertus (2015) found that the correlation between ANS acuity and more advanced mathematical abilities could not be explained by individual differences in children's inhibition. Using the modeling approach by DeWind et al. (2015) in a more sophisticated measurement, Starr and colleagues (2017) observed that four- and six-year-old children's number acuity and inhibition both contributed independently to general mathematical ability. Moreover, inhibition performance was not related to ANS acuity, which questions the idea that ANS acuity reflects inhibitory skills. The latter study, however, measured only one type of inhibitory skill, response suppression. Another form of inhibitory control is interference control (e.g., ignoring irrelevant features like in a Stroop task), which more closely mimics the processes needed in non-symbolic number comparison tasks. It has been demonstrated that both types of inhibitory control are differently related to non-symbolic number comparison performance in adults (Reynvoet et al., 2021).

Since the results on the role of inhibition are not entirely conclusive, more research is needed in order to find a more accurate and optimal ANS measurement method. In the meantime, it seems to be beneficial to measure and control for inhibitory processes (in particular interference control) when measuring ANS acuity (Fuhs & McNeil, 2013; Gilmore et al., 2013; Keller & Libertus, 2015). Note, however, that the reliability of the variables may considerably influence the regression results (see the Reliability section below). Importantly, inhibition scores notoriously show low reliability (Rouder et al., 2019), and many inhibition indices are gain scores (i.e., a difference of two conditions/scores such as congruent and incongruent trial reaction times), that are known to have lower reliability than its components' reliabilities in many circumstances (Draheim et al., 2019).

If non-numerical perceptual properties and inhibitory components are not identified and separated appropriately, the correlation between the biased ANS index and general mathematical ability may be either artificially increased or decreased compared to the true correlation. The observed correlation will not be the correlation of true ANS acuity and mathematical ability, but the correlation of an unknown mix of true ANS acuity, perceptual sensitivities, and inhibition, with mathematical ability. Depending on the loads of the components on the biased ANS index and the real correlation between those components and mathematical ability, the observed correlation could be either higher or lower than the real correlation between true ANS acuity and general mathematical ability. This means that, in research that does not control for non-numerical features and inhibition appropriately, the results cannot be interpreted validly because one cannot be sure which component(s) of the measured ANS index is/are responsible for the observed correlation.

An alternative to directly choosing an independent measure of inhibitory control and controlling for it when assessing the relation between ANS acuity and mathematical ability is to, where possible, adopt one of the analyses used by Gilmore et al. (2013). They examined whether the ANS/mathematics ability correlation they observed was driven by their task's (dot size and convex hull) congruent comparison trials, or the incongruent trials. A substantial relation between performance on incongruent trials and mathematics ability was found, but no such relation was present for the congruent trials. This cannot be explained by ceiling performance on the congruent trials. Based on this, Gilmore et al. concluded that the overall relation they observed was likely due to inhibitory control. This within-task approach avoids the difficulty of choosing an appropriate independent task for assessing inhibition with an adequate level of reliability (Rouder et al., 2019).

In addition to inhibition, various domain-general cognitive abilities appear to be intertwined with the Approximate Number System (ANS) in diverse manners. The existing literature presents inconclusive evidence regarding the influence of these domain-general skills on the relation between ANS acuity and mathematics, with studies finding unique contributions of the ANS to mathematical abilities (Coolen et al., 2023; Libertus et al., 2013a; Passolunghi et al., 2014), but other findings suggesting that the link between ANS acuity and mathematics ceases to exist when considering cognitive skills such as working memory and visuospatial skills (Caviola et al., 2020; Coolen et al., 2022; Costa et al., 2021). While a thorough review of domain-general contributions in ANS tasks is beyond the scope of the current work, they are worth mentioning as they should also be taken into consideration.

In sum, it is recommended to (1) model the contribution of non-numerical features and perceived dimensions (DeWind et al., 2015; Tomlinson et al., 2020), and (2) assess the role of inhibition (although it is an open question which methods and paradigms are optimal for this). Also, complete stimulus descriptions (including code used to generate them and the stimuli themselves if possible) should be provided for re-analysis of potentially confounding variables.

Operationalizing ANS acuity

Even in a simultaneously presented, non-symbolic number comparison task, ANS acuity can be calculated in various ways. These indices may considerably differ in terms of reliability and validity. Here, three frequently used ANS acuity indices are discussed: the Weber fraction, the ratio effect slope, and the mean correct responses (see references for studies using different indices in Chen & Li, 2014; Schneider et al., 2017). When applied in developmental studies, all of these indices have additional issues that researchers should consider.

Weber fraction

ANS acuity can be calculated with psychophysical methods. There are various ways to calculate the Weber fraction. One method that has become popular in the numerical cognition literature relies on signal detection theory (SDT) (Dehaene, 2007; Kingdom & Prins, 2010). The proportion of the correct responses is found with the formula $\frac{1}{2}erfc\left(\frac{|n_1-n_2|}{\sqrt{2}w\sqrt{n_1^2+n_2^2}}\right)$, where n_1 and n_2 are the to-be-compared values, w is the internal Weber fraction, and *erfc* is the so-called complementary error function. The formula can be extended with the *lapse rate* parameter: Participants sometimes respond incorrectly because of, for example, a lapse in attention to the task, or for other reasons that are independent of their ability to perform the numerical comparison task itself (Pica et al., 2004). In this context, the lapse rate parameter is the upper asymptote of the sigmoid curve, that is, an upper limitation of performance that reflects the general inability to pay attention to the task.

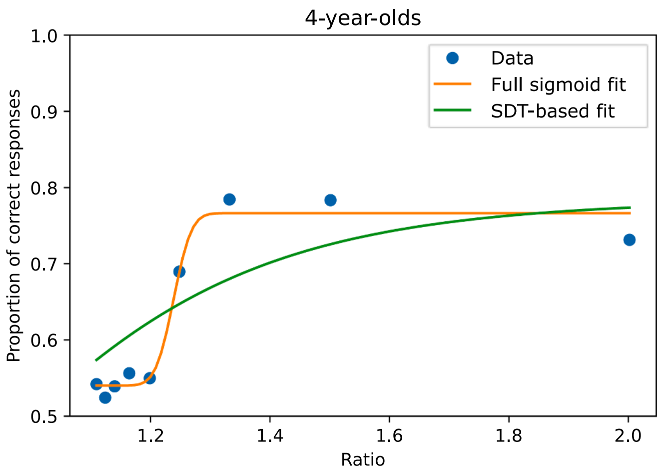

Fig. 1. Percent correct (y-axis) as a function of ratio (x-axis) in four-year-old children, where SDT-based fit (green line) and full sigmoid fit (orange line) are shown (redrawn after Figure 3 in Halberda & Feigenson, 2008). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Although this method has been used in many studies measuring ANS acuity in children, there are various issues when following these calculations. First, the response pattern observed in children may deviate from the pattern that the specific function expects (Halberda & Feigenson, 2008). The SDT-based function (see the green line in Fig. 1) in its default form cannot capture response patterns that remain at chance even as the stimulus ratios increase from the small to medium range. Moreover, because of the specific shape, the SDT-based function assumes accuracy increases gradually, the responses for the easy number pairs displaying a plateau are not well captured either. The difference between the SDT-based description and a more general sigmoid function (see the orange line in Fig. 1) leads to different Weber fraction estimates, and this difference is larger in younger children: The difference between the Weber fractions using the various methods could be as large as 0.2 in three-year-old children (see Table 5 in Halberda & Feigenson, 2008). Making things even more complicated, unlike in the SDT-based function, in the more general sigmoid model, the location instead of the scale parameter is used as the Weber fraction. Unfortunately, general psychophysical considerations cannot help determine which parameter should be chosen as the acuity index (Kingdom & Prins, 2010). Note that the method that considers numerical and non-numerical properties at the same time also relies on the SDT-based function (DeWind et al., 2015); therefore, the observed deviating responses compared to the SDT-based function could be an issue in this method as well. The discrepancy between the SDT-based function and a full sigmoid function is hard to resolve. Even a more general psychophysical consideration suggests that although SDT assumes an upper-half-sigmoid function performance (e.g., as the green line in Fig. 1), often a full sigmoid (e.g., the orange line in Fig. 1) is observed, and a logarithmic transformation can connect the two types of functions (Kingdom & Prins, 2010). Overall, despite the popularity of the SDT-based upper-half-sigmoid function in numerical cognition, psychophysical considerations do not provide unambiguous recommendations as to how performance should be modeled, and, while the SDT-based function fits adult non-symbolic number comparison performance appropriately, the same function may mischaracterize children's performance. Similarly, it is an open question whether scale or location parameters of the sigmoid functions should be preferred as ANS acuity indices. Unfortunately, to our knowledge, it is not yet known which psychophysical Weber fraction measurement method should be preferred in young children and how the discrepancies mentioned here should be resolved.

An SDT-based function without the lapse rate parameter may lead to additional problems. It was found that the Weber fraction calculated without the lapse rate parameter (i.e., assuming that performance reaches 100 % mean accuracy) strongly correlates with the lapse rate itself, that is, the assumed Weber fraction is, in fact, dominantly a lapse rate index (Chesney et al., 2015). This indicates that Weber fraction calculations that do not account for attention (i.e., without the lapse rate parameter) to the task may not be valid. Therefore, functions with a lapse rate parameter should be used (Chesney et al., 2015; Pica et al., 2004), particularly with populations - such as children - whose consistent ability to pay attention to the task may be in doubt.

Overall, while the Weber fraction could be an ideal solution because it reflects ANS acuity in a theoretically motivated way (in the sense that various parameters of a psychophysical process can be separated), it is still not clear which approach should be used to calculate Weber fractions (Halberda & Feigenson, 2008; Kingdom & Prins, 2010). Weber fractions calculated without consideration of lapse rate - such as via the popular SDT method - may be a biased index of ANS acuity. Therefore, the lapse rate should be considered when calculating the Weber fraction (Chesney et al., 2015; Halberda & Feigenson, 2008; Pica et al., 2004), and the same is recommended for the perceptual properties of the stimuli (DeWind et al., 2015). While the SDT-based method seems to be biased when applied to young children's data, it is not straightforward yet what function should be used to fit the data and which parameter should be applied as the ANS acuity (Halberda & Feigenson, 2008).

Ratio effect slope

According to the ANS model, in a comparison task, performance is proportional to the ratio of the to-be-compared values. As an alternative to calculating the Weber fraction, a number of researchers have instead attempted to capture individual differences in ANS acuity by finding the size or slope of the ratio or distance effect (see Inglis & Gilmore, 2014). Mean performance on numerical comparison tasks at two or more stimulus ratios is assessed. Then researchers use a linear regression to regress performance on the comparison ratio and find a slope. When only two stimulus ratios are used (or the ratios are grouped into two categories – small and large ratios), researchers may bypass the need for linear regression and simply find the difference in mean accuracy between the two ratio conditions. Alternatively, performance can be calculated as a function of distance, instead of the ratio; therefore, the distance effect slope can be used instead of the ratio effect slope.

Importantly, the relation between the ratio effect slope and the Weber fraction is not only non-linear but also non-monotonic (Chesney, 2018). This means that while, in some cases, a larger slope means a larger Weber fraction, in some other cases, a larger slope means a smaller Weber fraction. Obviously, this non-monotonic relation may decrease, abolish, or even reverse correlations between this measure of ANS acuity and other abilities or performance. This non-monotonic relation may also explain why the reliability of the distance effect slope can be lower than the reliability of the Weber fraction (Lindskog et al., 2013). A more detailed analysis revealed that the effect of this non-monotony depends on the Weber fraction range and the ratios used in the paradigm (Krajcsi, 2020). While small Weber fractions that are observable in adults can sometimes produce a relatively linear relation between the slope and the Weber fraction, measurements of children with larger Weber fractions may be influenced more strongly by this non-monotonic relation. Relatedly, the observed variability may also depend on these factors, leading to a smaller variance of the ratio effect and consequently lower reliability in children (find more details about the unintuitive relation of the relevant indices in Chesney, 2018; Krajcsi, 2020). This may be partly the explanation of why the ratio effect slope has lower reliability in children compared to adults (Inglis & Gilmore, 2014). Smaller ratio effect slope variability and lower reliability in children may also partly account for the result from different *meta*-analyses (Chen & Li, 2014; Fazio et al., 2014; Schneider et al., 2017) that the ratio effect slope correlates more weakly with mathematical abilities than the Weber fraction.

To summarize, the ratio/distance effect slope has a non-monotonic relation with the Weber fraction (Chesney, 2018). In children, its reliability may be low, and its use in correlational studies is severely limited (Krajcsi, 2020). Overall, the ratio/distance effect slope as an index of ANS acuity is not recommended for capturing ANS acuity in young children (Chesney, 2018; Dietrich et al., 2015; Krajcsi, 2020).

Mean accuracy

Mean accuracy is often calculated to assess individual differences in ANS acuity. One clear advantage of this method is its simplicity. Additionally, it has been demonstrated that accuracy often correlates with the Weber fraction as calculated with more sophisticated psychophysical methods (Fazio et al., 2014; Gilmore et al., 2013; Inglis & Gilmore, 2014).

An important disadvantage of this method, as opposed to the Weber fraction calculation, is that accuracy is sensitive to the applied ratios; that is, good performance can result from both high ANS acuity and easy-to-detect ratios. This also means that mean accuracy cannot be compared across studies if the ratios differ.

Another important limitation of the mean accuracy as an index of ANS acuity is that it is strongly influenced by the lapse rate parameter. A higher lapse rate causes lower mean accuracy, independent of ANS acuity. Consequently, mean accuracy is an unknown mixture of ANS acuity, and attentional processes (or anything else that may influence the lapse rate parameter). Similar to other biased indices of ANS acuity described above, the biased mean accuracy can yield higher or lower correlations between this biased mean accuracy and general mathematical ability compared to the real correlation between true ANS acuity and general mathematical ability. In young children, this issue could be quite serious: Three- and four-year-old children may show a high lapse rate, with only a 75 % correct response rate for large ratios that should be discriminable by their ANS (see Figure 3 in Halberda & Feigenson, 2008). This latter 75 % lapse rate means that in around 50 % of the trials, the numerical task is not performed. In young children, the lapse rate is quite high, and thus the role of the lapse rate could be even more detrimental to the validity of using mean accuracy as an index of ANS acuity, as would be the case for adults.

How heavily can the lapse rate influence the mean accuracy, and how heavily can it decrease the load of ANS acuity in that index? As mentioned above, in some studies (Gilmore et al., 2013; Inglis & Gilmore, 2014), Weber fraction and accuracy were correlated, which may hint that the lapse rate does not cause a substantial bias in the accuracy index. However, the Weber fraction can be calculated with different methods (see several considerations above), and the calculation may or may not consider the lapse rate. Chesney and her colleagues (2015) found that the lapse rate strongly correlates with the Weber fraction if the lapse rate is not considered a separate parameter in the Weber fraction calculation procedure ($r = 0.8$), but the correlation is 0.02 if the lapse rate is considered. This means that if the lapse rate is not considered as a separate parameter, the found Weber fraction may be strongly biased by the lapse rate. (Conversely, researchers should also include a sufficiently broad range of stimulus ratios in their tasks to determine that lapse rate calculations truly reflect lapse rate; for example, by demonstrating an asymptote on the sigmoid function as in Fig. 1. Otherwise, poor performance by participants with very low acuity may incorrectly be attributed to high lapse rates, thus artificially inflating such participants' acuity assessments.) In the studies above that showed correlations between the Weber fraction and mean accuracy, the lapse rate was not considered in the Weber fraction calculation. Because the Weber fraction could be strongly biased in those studies, the correlation between the Weber fraction estimates and mean accuracy rates may be largely accounted for by a shared correlation with the lapse rate rather than by a shared correlation with underlying ANS acuity. Further complicating matters, Chesney et al. (2015) also found that higher lapse rates predicted poorer mathematics performance. Thus, failing to account for the lapse rate might artificially inflate apparent correlations between ANS acuity and mathematical ability.

In practice, mean accuracy is often preferred over the Weber fraction because the reliability of the mean accuracy may be higher than the reliability of the Weber fraction. In other words, the mean accuracy is less noisy than the Weber fraction index. Relatedly, a sigmoid fit may not converge for some participants (e.g., see Libertus et al., 2011), while mean accuracy can always be calculated. Mathematically, this is understandable: mean accuracy tries to find a single parameter, while a sigmoid fit tries to find several parameters (usually 2–4 parameters, depending on the specific function), also assuming a specific shape of the data. Higher reliability and calculability are apparent advantages of the mean accuracy. However, this comes with a price: The mean accuracy index merges various indices of the sigmoid fit which have different meanings (as discussed above). In other words, the meaning and the validity of the mean accuracy metric are unclear. Therefore, as compared to Weber fraction estimates, the simplicity and greater reliability of mean accuracy indexes come at the cost of lower validity. Moreover, the fact that Weber fraction estimates can fail to converge with untypical response patterns can be viewed as essentially a built-in manipulation check. Generally, convergence fails when participants are not more accurate on easier trials. If participants are not more accurate on easier trials, then the task is not properly capturing ANS estimation (i.e., the trials are all too easy, all too hard, or the participant is not following the instructions). Such validity issues are flagged by convergence failures in Weber fraction estimations, but may go unnoticed if one is only considering mean accuracy.

Overall, the lapse rate may heavily influence the validity of mean accuracy as an index of ANS acuity. Until the role of lapse rate is clarified, this index should be treated with reservations. Also, the mean accuracy is an index of not only the ANS acuity and the lapse rate but also the task difficulty (i.e., number ratios). While Weber fraction calculation removes the role of the task difficulty and lapse rate, mean accuracy cannot; therefore, unless special care is given to the interpretation of the mean accuracy index, Weber fraction is preferred over mean accuracy as an index of ANS acuity.

Consider reliability

Many studies measuring ANS acuity in young children are designed to investigate the relation between ANS acuity and general mathematical ability via correlational methods. Reliability is essential in such studies because it limits the observable upper boundaries of correlations of ANS acuity with other mathematics-related measures, which are usually the focus of the researchers' interest. More generally, the (un)reliability of a task attenuates these correlations (Spearman, 1904): Low reliability means relatively larger noise in the variable (which can also be considered as additional unexplained variability of the variables), which in turn makes the observed correlation lower than the true correlation. This attenuation has several consequences that are critical in the interpretation of the results of the above correlational studies. First, since the observed correlation is lower than the true correlation, the power is smaller, and some non-zero correlations cannot be significant with a sample size that otherwise could be significant with indices showing better reliability (e.g., Hedge et al., 2018). Second, the effect sizes are systematically underestimated, and the true correlation cannot be observed or estimated. This can be critical in some cases. For example, Piazza (2010) proposed that ANS acuity alone can be responsible for the development of subset-knowers' number knowledge. This means that the true correlation between number knowledge and the appropriate transformation of the ANS acuity should be 1 (see a more complete mathematical description of this possible connection in Krajcsi et al., 2023). However, this prediction cannot be tested if the reliability and, consequently, the attenuation are not known. Third, since the reliabilities of several task indices can be different, the different attenuation effects on the correlations mean that the correlations cannot be compared. For example, a popular question of several *meta*-analyses (Chen & Li, 2014; Schneider et al., 2017) has been whether symbolic comparison skills are more/less related to mathematical performance than non-symbolic comparison performance. However, the observed correlation can differ not only because the true correlations are different but also because the reliabilities of those measurements may be different (see our example above about the different reliability of the ANS indices). Overall, papers reporting correlational analyses should always consider that the observed correlations may be smaller than the true correlations because of reliability-based attenuation. Note again that these considerations are relevant only in correlational studies; however, the reliability as measured with common tools (such as test–retest correlation) does not directly influence studies that compare means (see the complex relation of reliability and power of central tendency comparison tests in Parsons, 2018; Zimmerman & Zumbo, 2015).

The attenuation-related issues can be handled if the reliabilities of the indices are known. Note, that in a correlation, the reliability of both variables is relevant; in our examples, the reliability of both the ANS acuity and the mathematics performance should be considered when interpreting the correlation to differentiate between low correlation because of conceptual reasons and low correlation because of reliability issues. A favorable trend in the last decade has been that several recent studies (Braham & Libertus, 2017; Chesney et al., 2015; Clayton et al., 2015; DeWind et al., 2015; DeWind & Brannon, 2012, 2016; Elliott et al., 2023; Gilmore et al., 2011, 2014; Inglis & Gilmore, 2014; Libertus et al., 2013a, 2013b; Lindskog et al., 2013; Maloney et al., 2010; Price et al., 2012; Sasanguie et al., 2011; Shusterman et al., 2016; Smets et al., 2014) have reported the reliabilities of a variety of indices. While many of the indices show high reliabilities, the overall picture is rather mixed. Importantly, the already reported reliabilities cannot always be informative for later studies for several reasons. (a) Cognitive scientists sometimes ignore that reliabilities of measures of a paradigm (e.g., non-symbolic comparison) will be different if the specific parameters of those methods are different (e.g., differences in number of trials, stimulus ratios, controls for continuous extent). For example, the reliability of a task will typically decrease if fewer trials are used (Dietrich et al., 2015; Krajcsi, in preparation; Lindskog et al., 2013; Rouder & Haaf, 2019). It is also clear that paradigms with different perceptual control methods may have different reliability. Additionally, reliability is specific to given indices; for example, the reliability of one index of ANS acuity cannot be assumed to be indicative of the reliability of a different index of ANS acuity. While ANS acuity can be accessed via different methods and tasks (see above), reported reliabilities are relevant only for the specific task *and* the specific index calculation method at hand. (b) Reliability is mostly a relative measure of the noise compared to the variability of the sample; therefore, reliability depends not only on the task but also on the variance of the participants. For example, a more heterogeneous sample leads to higher assessed reliability. This also means that, when a sample with larger variability is chosen (e.g., a sample with a larger age range), the assessed reliability may be better, even if the noise in the task is constant. (c) We do not know much about the fluctuation of ANS acuity in children over time (conceptually spoken state and trait aspects of ANS). For instance, if there is a large day-to-day fluctuation, split-half reliability would be larger than test–retest reliability. (d) The data of children can be noisier than the data of adults, and younger children may show larger noise than older children (Inglis & Gilmore, 2014). Noise decreases reliability; therefore, different age groups may show different reliability only because of the difference in the overall noise in the measurements. To sum up, because reliability may depend both on the task parameters that often vary between studies and on the samples themselves, reliabilities reported in one study cannot always be directly applied to the results of other studies. Therefore, it is recommended to report reliability in all correlational studies where the task(s) or the sample is changed compared to previous studies that have reported reliability. Reliability can be reported as test–retest correlation (if the task was measured twice), even–odd split-half reliability with Spearman-Brown correction (when the task was measured only once), or with other indices that are recommended by the methodological literature (Pronk et al., 2022). This recommendation extends not just to indices of ANS acuity but to all variables that are relevant in the correlations.

For young children, the length of a testing session is highly limited, which means that researchers must keep the number of trials low. However, fewer trials make the task less reliable; while the specific number of trials is not the only factor that determines the reliability, often hundreds of trials are recommended (Kingdom & Prins, 2010; Lindskog et al., 2013). Because of the high variability in the responses, the Weber fraction cannot always be found, for example, with 35 trials for four- and five-year-old children (Odic et al., 2014). Additionally, a high lapse rate (e.g., around 75 % maximum performance of three- and four-year-old children, indicating 50 % of trials when the numerical task is not performed; Halberda & Feigenson, 2008) means that a large proportion of trials is only noise, making the reliability low. This constraint makes it even harder to reach an appropriate level of reliability for an ANS acuity index in children. Importantly, some methods may decrease the number of necessary trials, such as the adaptive measurement procedure (Lindskog et al., 2013). Considering both factors, in some cases, it is not easy to find an appropriate compromise, or it may be impossible to find an appropriate solution: For some indices, it is possible that no reliable measurement can be designed, and it may be impossible to measure the individual differences of some abilities appropriately (see a similar failure even in adults in Rouder et al., 2019). In some other cases, a reliable index requires so many trials (e.g., 13,400 trials for an appropriate priming distance effect index in adults; Krajcsi & Szűcs, 2022) that it cannot be achieved with child participants. Note that it is not only the reliability of the indices that may be compromised by the small number of trials but also the precision of the reliability measurements (i.e., the reliability index may be noisy or, in other words, may not be reliable). This latter issue makes the evaluation of the index in young children rather difficult.

To summarize, since reliability can attenuate correlations, it is always essential to consider reliability in correlational studies. Because reliability depends on the task parameters, the population, and the indices, formerly reported reliabilities of a paradigm are not always sufficient to be relied upon for new studies; therefore, reliabilities of the relevant indices should be reported for each data set. However, in many cases, gaining appropriate reliability (or specifying the reliability at all) can be challenging, or even impossible. For this reason, it is essential that reports of low reliabilities should also be made available in the literature because it is essential information for other researchers. From the viewpoint of publication and editorial decision-making, this means that because reliability is a critical detail of a study, high reliability in itself should not be a precondition for accepting a manuscript. We view this as absolutely critical because if only studies with high reliabilities are published (Hussey et al., 2023), we would get a misleading picture in the literature in the same way as we had, and still have, in the replication crisis where null effects were not published. Therefore, we would overestimate the reliabilities of the tasks we are using, when low reliability studies were not published. However, for both high and low reliabilities, reliability needs to be addressed and not just neglected. From the viewpoint of positive results, if appropriate reliability cannot be reached, then the limitations of the measurements should be recognized and alternative approaches should be found to test the relevant theoretical questions.

Conclusions

Measuring ANS acuity is essential in resolving theoretical debates, particularly about the development of numerical abilities and potentially in clinical works. Therefore, investigators need to be aware of the pros and cons of various methods used to assess ANS acuity, and particularly how known methodological issues may compromise the assessment of ANS acuity in young children. Here, we reviewed several methodological issues in popular ANS acuity metrics, with a focus on non-symbolic number comparison tasks (see Table 1). Based on our review of the literature, we make the following methodological recommendations: (1) Ensure that even young children understand the task; (2) Account for the non-numerical perceptual features of the stimuli in the analysis; (3) Account for

Table 1
Summary of the issues and potential solutions in measuring the ANS in preschoolers. See more details in the appropriate sections of this paper.

Issues	Solutions and challenges
Understanding the task	
Children may not understand the “choose more/fewer dots” instruction (Lindskog & Simms, 2021; Negen & Sarnecka, 2015).	Alternative task versions, e.g., training trials (Negen & Sarnecka, 2015) or nonverbal demonstration (Lindskog & Simms, 2021) without feedback (Odic et al., 2014), ensure that children understand the task.
Non-numerical features of the stimuli	
Non-numerical perceptual properties influence the measured ANS acuity (Gebuis & Reynvoet, 2012).	Smarter controls for more unconfounded effects (e.g., Gebuis & Reynvoet, 2011). Analyses that consider and separate the various sources of the decision (DeWind et al., 2015). For future reanalyses, consistently and thoroughly report parameters of all features of the stimuli (e.g., descriptives of dot sizes, cumulative surface areas, and spacing).
Inhibitory components may play a role (Fuhs & McNeil, 2013; Gilmore et al., 2013; but see Elliott et al., 2019; Keller & Libertus, 2015).	The role of inhibitory control should be assessed and considered in the analysis (Elliott et al., 2019; Fuhs & McNeil, 2013; Gilmore et al., 2013; Keller & Libertus, 2015).
Indices of ANS Acuity	
Some Weber fraction calculation methods may be biased (Halberda & Feigenson, 2008).	The lapse rate (Chesney et al., 2015) and perceptual properties (DeWind et al., 2015) should be considered. Currently, it is unknown which function is preferred.
Slope relates to the Weber fraction non-monotonically (Chesney, 2018; Krajcsi, 2020).	Ratio (or distance) effect slopes are not recommended as an ANS acuity index.
Mean accuracy may be biased by the lapse rate (see various methods in Chesney et al., 2015; Gilmore et al., 2013; Inglis & Gilmore, 2014).	Currently, it is unknown how strongly the mean accuracy is biased.
Reliability	
Reliability may be low, which influences the measured correlation (Spearman, 1904).	Reliability should be reported and considered for the ANS index and also all other target indices.
Reliability may be a more serious issue in young children compared to older children or adults.	Reliability can be improved with more trials, adaptive procedures (Lindskog et al., 2013), and a wider age range. However, increased numbers of trials may not be practical with young children.
In some cases, it may be impossible to measure an index reliably (Rouder et al., 2019).	Alternative research strategy/design/paradigm/index should be found.

inhibition abilities in the analysis; (4) Consider lapse rate when calculating an ANS acuity index; (5) Do not use the ratio/distance effect slope an ANS index; (6) Report reliability and consider its role in the analyses. There are several issues for which no appropriate solution or recommendation is available at the moment, such as which instruction variants are optimal, what is an optimal way to consider the role of inhibition, which Weber fraction calculation method is preferred, and how to provide appropriate reliability.

In recent years, there has been important progress in improving the methods for measuring the acuity of the ANS, and, at the same time, new issues have been raised. It is essential that future studies consider these issues and rely on the available solutions. Also, based on these considerations, many previous works should be reinterpreted in order to reach more valid conclusions. In addition, based on these considerations, it is also straightforward to (conceptually) replicate or reanalyze former studies to find more valid and reliable results. Finally, additional methodological works are needed to find solutions for the methodological issues not yet resolved and to identify and resolve additional issues that may compromise the interpretation of the measured ANS acuity. We hope researchers find this paper useful in designing future investigations and filling the gaps in our understanding of the best indices of ANS acuity for young children.

Funding

Attila Krajcsi was supported by the CELSA Research Fund (project no. CELSA/19/011) and by the National Research, Development and Innovation Fund of Hungary (project no. 132165). Krzysztof Cipora, Camilla Gilmore, Matthew Inglis and Victoria Simms were supported by the UKRI Economic and Social Research Council (grant number ES/W002914/1). Melissa Libertus was supported by a Scholar Award from the James S. McDonnell Foundation. Melissa Libertus' contribution to the present paper also benefited from synergistic activities on the ManyNumbers project funded by the National Science Foundation of the United States (DRL 2201960).

CRediT authorship contribution statement

Attila Krajcsi: Conceptualization, Writing – original draft, Writing – review & editing. **Dana Chesney:** Writing – review & editing. **Krzysztof Cipora:** Writing – review & editing. **Ilse Coolen:** Writing – review & editing. **Camilla Gilmore:** Writing – review & editing. **Matthew Inglis:** Writing – review & editing. **Melissa Libertus:** Writing – review & editing. **Hans-Christoph Nuerk:** Writing – review & editing. **Victoria Simms:** Writing – review & editing. **Bert Reynvoet:** Conceptualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

Aulet, L. S., & Lourenco, S. F. (2021). The relative salience of numerical and non-numerical dimensions shifts over development: A re-analysis of Tomlinson, DeWind, and Brannon (2020). *Cognition, 210*, Article 104610. https://doi.org/10.1016/j.cognition.2021.104610

Braham, E. J., & Libertus, M. E. (2017). Intergenerational associations in numerical approximation and mathematical abilities. *Developmental Science, 20*(5), e12436.

Braham, E. J., & Libertus, M. E. (2018). When approximate number acuity predicts math performance: The moderating role of math anxiety. *PLOS ONE, 13*(5), e0195696.

Brannon, E. M., & Merritt, D. J. (2011). Chapter 14—Evolutionary Foundations of the Approximate Number System. In S. Dehaene, & E. M. Brannon (Eds.), *Space, Time and Number in the Brain* (pp. 207–224). Academic Press. https://doi.org/10.1016/B978-0-12-385948-8.00014-1.

Butterworth, B. (2010). Foundational numerical capacities and the origins of dyscalculia. *Trends in Cognitive Sciences, 14*(12), 534–541. https://doi.org/16/j.tics.2010.09.007.

Carey, S. (2009). *The origin of concepts* (1st ed.). USA: Oxford University Press.

Carey, S., & Barner, D. (2019). Ontogenetic origins of human integer representations. *Trends in Cognitive Sciences, 23*(10), 823–835. https://doi.org/10.1016/j.tics.2019.07.004

Caviola, S., Colling, L. J., Mammarella, I. C., & Szűcs, D. (2020). Predictors of mathematics in primary school: Magnitude comparison, verbal and spatial working memory measures. *Developmental Science, 23*(6), e12957.

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

Chesney, D. (2018). Numerical distance effect size is a poor metric of approximate number system acuity. *Attention, Perception, & Psychophysics, 80*(5), 1057–1063. https://doi.org/10.3758/s13414-018-1515-x

Chesney, D., Bjalkebring, P., & Peters, E. (2015). How to estimate how well people estimate: Evaluating measures of individual differences in the approximate number system. *Attention, Perception, & Psychophysics, 77*(8), 2781–2802. https://doi.org/10.3758/s13414-015-0974-6

Chesney, D., Shoots-Reinhard, B., & Peters, E. (2021). The Causal Impact of Objective Numeracy on Judgments: Improving Numeracy via Symbolic and Non-Symbolic Approximate Arithmetic Training Yields More Consistent Risk Judgments. *Journal of Numerical Cognition, 7*(3), 351–367. https://doi.org/10.5964/jnc.6925

Clayton, S., & Gilmore, C. (2015). Inhibition in dot comparison tasks. *ZDM, 47*(5), 759–770. https://doi.org/10.1007/s11858-014-0655-2

Clayton, S., Gilmore, C., & Inglis, M. (2015). Dot comparison stimuli are not all alike: The effect of different visual controls on ANS measurement. *Acta Psychologica, 161*, 177–184. https://doi.org/10.1016/j.actpsy.2015.09.007

Coolen, I. E. J. I., Omont-Lescieux, S., & Knops, A. (2023). *Now you see it, now you don't—Cognitive skills and their contributions to mathematics across early development* (1). *6*(1), Article 1. https://doi.org/10.5334/joc.309.

Coolen, I. E. J. I., Riggs, K. J., Bugler, M., & Castronovo, J. (2022). The approximate number system and mathematics achievement: It's complicated. A thorough investigation of different ANS measures and executive functions in mathematics achievement in children. *Journal of Cognitive Psychology*, 1–23. https://doi.org/10.1080/20445911.2022.2044338

Costa, H. M., Partanen, P., & Herwegen, J. V. (2021). *The role of Working Memory, Processing Speed and Approximate Number System abilities in low maths achievement among preschoolers*. PsyArXiv. https://doi.org/10.31234/osf.io/b6p8d.

Defever, E., Reynvoet, B., & Gebuis, T. (2013). Task- and age-dependent effects of visual stimulus properties on children's explicit numerosity judgments. *Journal of Experimental Child Psychology, 116*(2), 216–233. https://doi.org/10.1016/j.jecp.2013.04.006

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*, 1–42. https://doi.org/10.1016/0010-0277(92)90049-N

Dehaene, S. (2007). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In P. Haggard, Y. Rossetti, & M. Kawato (Eds.), *Sensorimotor foundations of higher cognition* (Vol. XXII, pp. 527–574). Harvard University Press.

DeWind, N. K., Adams, G. K., Platt, M. L., & Brannon, E. M. (2015). Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition, 142*, 247–265. https://doi.org/10.1016/j.cognition.2015.05.016

DeWind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: Effects of feedback and training. *Frontiers in Human Neuroscience, 6*, 68. https://doi.org/10.3389/fnhum.2012.00068

DeWind, N. K., & Brannon, E. M. (2016). Significant inter-test reliability across approximate number system assessments. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.00310

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*. https://doi.org/10.3389/fpsyg.2015.00295

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin, 145*(5), 508–535. https://doi.org/10.1037/bul0000192

Elliott, L., Feigenson, L., Halberda, J., & Libertus, M. E. (2019). Bidirectional, longitudinal associations between math ability and approximate number system precision in childhood. *Journal of Cognition and Development, 20*(1), 56–74. https://doi.org/10.1080/15248372.2018.1551218

Elliott, L., Votruba-Drzal, E., Miller, P., Libertus, M. E., & Bachman, H. J. (2023). Unpacking the home numeracy environment: Examining dimensions of number activities in early childhood. *Early Childhood Research Quarterly, 62*, 129–138. https://doi.org/10.1016/j.ecresq.2022.08.002

Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology, 123*, 53–72. https://doi.org/10.1016/j.jecp.2014.01.013

Feigenson, L., Dehaene, S., & Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*, 307–314.

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

Gebuis, T., & Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behavior Research Methods, 43*(4), 981–986. https://doi.org/10.3758/s13428-011-0097-5

Gebuis, T., & Reynvoet, B. (2012). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General, 141*(4), 642–648. https://doi.org/10.1037/a0026218

Gelman, R. (1972). Logical capacity of very young children: number invariance rules. *Child Development, 43*(1), 75–90. https://doi.org/10.2307/1127873

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., Simms, V., & Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLOS ONE, 8*(6), e67374.

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., & 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., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*(7213), 665–668. https://doi.org/10.1038/nature07246

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Hurewitz, F., Gelman, R., & Schnitzer, B. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences, 103*(51), 19599–19604. https://doi.org/10.1073/pnas.0609485103

Hussey, I., Alsalti, T., Bosco, F., Elson, M., & Arslan, R. C. (2023, February 3). An aberrant abundance of Cronbach's alpha values at .70. https://doi.org/10.31234/osf.io/dm8xn.

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

Keller, L., & Libertus, M. (2015). Inhibitory control may not explain the link between approximation and math abilities in kindergarteners from middle class families. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00685

Kingdom, F. A. A., & Prins, N. (2010). *Psychophysics: A practical introduction*. Academic.

Krajcsi, A. (2020). Ratio effect slope can sometimes be an appropriate metric of the approximate number system sensitivity. *Attention, Perception, & Psychophysics, 82*(4), 2165–2176. https://doi.org/10.3758/s13414-019-01939-6

Krajcsi, A. (in preparation). *Estimated reliability based on diffusion model parameters*.

Krajcsi, A., Fedele, M., & Reynvoet, B. (2023). The approximate number system cannot be the leading factor in the acquisition of the first symbolic numbers. *Cognitive Development, 65*, Article 101285. https://doi.org/10.1016/j.cogdev.2022.101285

Krajcsi, A., Kojouharova, P., & Lengyel, G. (2022). Processing Symbolic Numbers: The Example of Distance and Size Effects. In J. Gervain, G. Csibra, & K. Kovács (Eds.), *A Life in Cognition: Studies in Cognitive Science in Honor of Csaba Pléh* (pp. 379–394). Springer International Publishing. https://doi.org/10.1007/978-3-030-66175-5_27.

Krajcsi, A., & Szűcs, T. (2022). Symbolic number comparison and number priming do not rely on the same mechanism. *Psychonomic Bulletin & Review, 29*(5), 1969–1977. https://doi.org/10.3758/s13423-022-02108-x

Leibovich, T., & Ansari, D. (2016). The symbol-grounding problem in numerical cognition: A review of theory, evidence, and outstanding questions. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 70*(1), 12–23. https://doi.org/10.1037/cep0000070

Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2017). From "sense of number" to "sense of magnitude": The role of continuous magnitudes in numerical cognition. *Behavioral and Brain Sciences, 40*. https://doi.org/10.1017/S0140525X16000960

Libertus, M. E., & Brannon, E. M. (2010). Stable individual differences in number discrimination in infancy. *Developmental Science, 13*(6), 900–906. https://doi.org/10.1111/j.1467-7687.2009.00948.x

Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science, 14*(6), 1292–1300. https://doi.org/10.1111/j.1467-7687.2011.01080.x

Libertus, M. E., Feigenson, L., & Halberda, J. (2013a). Is approximate number precision a stable predictor of math ability? *Learning and Individual Differences, 25*, 126–133. https://doi.org/10.1016/j.lindif.2013.02.001

Libertus, M. E., Feigenson, L., & Halberda, J. (2013b). Numerical approximation abilities correlate with and predict informal but not formal mathematics abilities. *Journal of Experimental Child Psychology, 116*(4), 829–838. https://doi.org/10.1016/j.jecp.2013.08.003

Lindskog, M., & Simms, V. (2021). *Measuring toddlers non-symbolic number processing*. https://doi.org/10.31234/osf.io/v835g.

Lindskog, M., Winman, A., Juslin, P., & Poom, L. (2013). Measuring acuity of the approximate number system reliably and validly: The evaluation of an adaptive test procedure. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00510

Lyons, I. M., Ansari, D., & Beilock, S. L. (2015). Qualitatively different coding of symbolic and nonsymbolic numbers in the human brain. *Human Brain Mapping, 36*(2), 475–488. https://doi.org/10.1002/hbm.22641

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*(2), 154–161. https://doi.org/10.1016/j.actpsy.2010.01.006

Mix, K. S., Levine, S. C., & Huttenlocher, J. (1997). Numerical abstraction in infants: Another look. *Developmental Psychology, 33*(3), 423–428. https://doi.org/10.1037/0012-1649.33.3.423

Mou, Y., Xiao, H., Zhang, B., Jiang, Y., & Wang, X. (2023). Are they equivalent? An examination of task variants of approximate number comparison. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02223-0.

Moyer, R. S., & Landauer, T. K. (1967). Time required for Judgements of Numerical Inequality. *Nature, 215*(5109), 1519–1520. https://doi.org/10.1038/2151519a0

Navarro, M. G., Braham, E. J., & Libertus, M. E. (2018). Intergenerational associations of the approximate number system in toddlers and their parents. *British Journal of Developmental Psychology, 36*(4), 521–539. https://doi.org/10.1111/bjdp.12234

Negen, J., & Sarnecka, B. W. (2015). Is there really a link between exact-number knowledge and approximate number system acuity in young children? *British Journal of Developmental Psychology, 33*(1), 92–105. https://doi.org/10.1111/bjdp.12071

Nieder, A., & Dehaene, S. (2009). Representation of Number in the Brain. *Annual Review of Neuroscience, 32*(1), 185–208. https://doi.org/10.1146/annurev.neuro.051508.135550

Odic, D., Hock, H., & Halberda, J. (2014). Hysteresis affects approximate number discrimination in young children. *Journal of Experimental Psychology. General, 143*(1), 255–265. https://doi.org/10.1037/a0030825

Odic, D., Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Developmental change in the acuity of approximate number and area representations. *Developmental Psychology, 49*(6), 1103–1112. https://doi.org/10.1037/a0029472

Odic, D., & Starr, A. (2018). An Introduction to the Approximate Number System. *Child Development Perspectives, 12*(4), 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*(10), 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*(1), 188–200. https://doi.org/10.1016/j.cognition.2014.06.011

Parrish, A. E., & Beran, M. J. (2022). Approximate Number System (ANS). In J. Vonk, & T. K. Shackelford (Eds.), *Encyclopedia of Animal Cognition and Behavior* (pp. 381–386). Springer International Publishing. https://doi.org/10.1007/978-3-319-55065-7_1068.

Parsons, S. (2018). *Visualising two approaches to explore reliability-power relationships*. PsyArXiv. https://doi.org/10.31234/osf.io/qh5mf.

Passolunghi, M. C., Cargnelutti, E., & Pastore, M. (2014). The contribution of general cognitive abilities and approximate number system to early mathematics. *The British Journal of Educational Psychology, 84*(Pt 4), 631–649. https://doi.org/10.1111/bjep.12054

Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences, 14*(12), 542–551. https://doi.org/10.1016/j.tics.2010.09.008

Piazza, M., De Feo, V., Panzeri, S., & Dehaene, S. (2018). Learning to focus on number. *Cognition, 181*, 35–45. https://doi.org/10.1016/j.cognition.2018.07.011

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and Approximate Arithmetic in an Amazonian Indigene Group. *Science, 306*(5695), 499–503. https://doi.org/10.1126/science.1102085

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*(1), 50–57. https://doi.org/10.1016/j.actpsy.2012.02.008

Pronk, T., Molenaar, D., Wiers, R. W., & Murre, J. (2022). Methods to split cognitive task data for estimating split-half reliability: A comprehensive review and systematic assessment. *Psychonomic Bulletin & Review, 29*(1), 44–54. https://doi.org/10.3758/s13423-021-01948-3

Reynvoet, B., Ribner, A. D., Elliott, L., Steenkiste, M. V., Sasanguie, D., & Libertus, M. E. (2021). Making Sense of the Relation Between Number Sense and Math. *Journal of Numerical Cognition, 7*(3), 308–327. https://doi.org/10.5964/jnc.6059

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review, 26*(2), 452–467. https://doi.org/10.3758/s13423-018-1558-y

Rouder, J. N., Kumar, A., & Haaf, J. M. (2019). *Why Most Studies of Individual Differences With Inhibition Tasks Are Bound To Fail* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/3cjr5.

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*(1), 73–80. https://doi.org/10.1016/j.actpsy.2010.10.004

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science, 20*(3). 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*(4), e0153072.

Slusser, E. B., & Sarnecka, B. W. (2011). Find the picture of eight turtles: A link between children's counting and their knowledge of number word semantics. *Journal of Experimental Child Psychology, 110*(1), 38–51. https://doi.org/10.1016/j.jecp.2011.03.006

Smets, K., Gebuis, T., Defever, E., & Reynvoet, B. (2014). Concurrent validity of approximate number sense tasks in adults and children. *Acta Psychologica, 150*, 120–128. https://doi.org/10.1016/j.actpsy.2014.05.001

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15*(1), 72–101. https://doi.org/10.2307/1412159

Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition, 36*(2), 97–127.

Starr, A., DeWind, N. K., & Brannon, E. M. (2017). The contributions of numerical acuity and non-numerical stimulus features to the development of the number sense and symbolic math achievement. *Cognition, 168*, 222–233. https://doi.org/10.1016/j.cognition.2017.07.004

Szkudlarek, E., & Brannon, E. M. (2017). Does the Approximate Number System Serve as a Foundation for Symbolic Mathematics? *Language Learning and Development, 13*(2), 171–190. https://doi.org/10.1080/15475441.2016.1263573

Tomlinson, R. C., DeWind, N. K., & Brannon, E. M. (2020). Number sense biases children's area judgments. *Cognition, 204*, Article 104352. https://doi.org/10.1016/j.cognition.2020.104352

Wynn, K. (1990). Children's understanding of counting. *Cognition, 36*(2), 155–193. https://doi.org/10.1016/0010-0277(90)90003-3

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology, 24*(2), 220–251. https://doi.org/10.1016/0010-0285(92)90008-P

Xenidou-Dervou, I., De Smedt, B., van der Schoot, M., & van Lieshout, E. C. D. M. (2013). Individual differences in kindergarten math achievement: The integrative roles of approximation skills and working memory. *Learning and Individual Differences, 28*, 119–129. https://doi.org/10.1016/j.lindif.2013.09.012

Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science, 8*(1), 88–101. https://doi.org/10.1111/j.1467-7687.2005.00395.x

Zimmerman, D., & Zumbo, B. (2015). Resolving the Issue of How Reliability is Related to Statistical Power: Adhering to Mathematical Definitions. *Journal of Modern Applied Statistical Methods, 14*(2). https://doi.org/10.22237/jmasm/1446350640.