# Comparing Numerical Comparison Tasks: A Meta-Analysis of the Variability of the Weber Fraction Relative to the Generation Algorithm

Mathieu Guillaume[1]* and Amandine Van Rinsveld[2]

[1] Cognitive Science and Assessment Institute (COSA), University of Luxembourg, Luxembourg, Luxembourg, [2] Centre for Research in Cognitive Neuroscience (CRCN), Université Libre de Bruxelles, Brussels, Belgium

Since more than 15 years, researchers have been expressing their interest in evaluating the Approximate Number System (ANS) and its potential influence on cognitive skills involving number processing, such as arithmetic. Although many studies reported significant and predictive relations between ANS and arithmetic abilities, there has recently been an increasing amount of published data that failed to replicate such relationship. Inconsistencies lead many researchers to question the validity of the assessment of the ANS itself. In the current meta-analysis of over 68 experimental studies published between 2004 and 2017, we show that the mean value of the Weber fraction (w), the minimal amount of change in magnitude to detect a difference, is very heterogeneous across the literature. Within young adults, w might range from <10 to more than 60, which is critical for its validity for research and diagnostic purposes. We illustrate here the concern that different methods controlling for non-numerical dimensions lead to substantially variable performance. Nevertheless, studies that referred to the exact same method (e.g., Panamath) showed high consistency among them, which is reassuring. We are thus encouraging researchers only to compare what is comparable and to avoid considering the Weber fraction as an abstract parameter independent from the context. Eventually, we observed that all reported correlation coefficients between the value of w and general accuracy were very high. Such result calls into question the relevance of computing and reporting at all the Weber fraction. We are thus in disfavor of the systematic use of the Weber fraction, to discourage any temptation to compare given data to some values of w reported from different tasks and generation algorithms.

**Keywords: Approximate Number System, number sense, meta-analysis, methodology, Weber fraction**

Over 20 years ago, Dehaene made the hypothesis that Humans possess a *Number Sense*, a biologically determined ability that allows us to represent and manipulate large numerical quantities (Dehaene, 1997). This numerical intuition is largely considered as relying on a cognitive system specifically dedicated to number processing called the *Approximate Number System* (ANS, Feigenson et al., 2004; see also Núñez, 2017; for an interesting terminological criticism). The crucial property of such cognitive system is the scalar variability of numerical approximations: numerical estimates of larger quantities are indeed more variable (Platt and Johnson, 1971; Gallistel and Gelman, 2000). Accordingly, the acuity of numerical discriminative processes handling two amounts is not absolute, but relative to the numerical ratio between the considered quantities (i.e., distinguishing 10 from 20 elements is easier than distinguishing 110 from 120 items). Mental number representations were thus hypothesized to go through a logarithmic compression following the Weber-Fechner law (Dehaene, 2003: but see Cantlon et al., 2009; Cicchini et al., 2014; and Piantadosi, 2016).

In order to assess these logarithmic representations, Piazza et al. (2004) were among the first to characterize performance (as well as brain activity) in a numerical discrimination task with the help of a measure directly related to the Weber-Fechner law, the Weber fraction. The Weber fraction is the ratio between the amount just noticeably different from a given magnitude, and the magnitude itself (*w*, see Stevens, 1957; Van Oeffelen and Vos, 1982). From a psychophysical perspective, the Weber fraction can be defined as the noise constant-proportionality parameter fitting the discrimination behavior during a numerical comparison task (see Barth et al., 2006, Appendix B). As a constant scaling ratio, the Weber fraction has the advantage of explicitly depicting the scalar variability across mental representations, which might fluctuate between individuals (see Whalen et al., 1999). More critically to the purpose of the current meta-analysis, this *w* parameter was heavily popularized in the literature as a direct measure of specific numerical quantity processes by some influential studies (e.g., Pica et al., 2004; Piazza et al., 2004). Subsequently, *w* was widely investigated as an individual property that is only subject to significant developmental changes across the lifespan (Halberda et al., 2012) and to refinement through formal instruction (Piazza et al., 2013). For a given age within a given population, *w* was thus considered as a stable predictor of more complex numerical processing such as math ability (Halberda et al., 2008), as well as a crucial clinical predictor of Mathematical Learning Disability (e.g., Mazzocco et al., 2011).

However, some authors recently questioned the stability of the Weber fraction. Due to the substantial amount of studies that were conducted following Halberda et al. study (2008), there were indeed many reports of failure in observing significant relationship between *w* and math ability (e.g., Price et al., 2012; Gilmore et al., 2013; Sasanguie et al., 2013). This raised some theoretical concerns (e.g., Gebuis et al., 2016; Leibovich et al., 2016; Núñez, 2017), as well as many methodological issues (see Dietrich et al., 2015b; for a review). Among these issues, many studies showed that the assessment of ANS acuity, and the measure of *w* itself, are not independent of interference from low level visual cues that are intrinsically confounded with

numerical quantities, and they revealed that *w* is nor consistent nor reliable across different tasks (Clayton et al., 2015, 2018; Bugden and Ansari, 2016; Guillaume et al., 2016). Some authors subsequently argued that the procedure used to generate visual arrays substantially influence participant behavior, and therefore the evaluation of w (Inglis and Gilmore, 2014; Clayton et al., 2015; Smets et al., 2015, 2016). In other words, *the Weber* fraction does not seem to be a stable psychophysiological parameter devoid of context; *w* can in fact be variable within one subject as a function of the task and the stimulus properties (but see Julio-Costa et al., 2015; DeWind and Brannon, 2016; for contradicting evidence).

Inglis and Gilmore (2014) went further by experimentally assessing the validity and the reliability of the Weber fraction in comparison to other measures of ANS acuity. Critically, they claimed that *w* was problematic for many reasons: its distribution was not normal but right-skewed, its test-retest reliability was poorer than every other measure of ANS acuity, and more fundamentally, its value was still affected by the way low level visual cues were manipulated in the task. These results do not support the view that the Weber fraction is an invariable psychophysiological parameter devoid of context. In addition, the authors reported that *w* highly correlated with overall accuracy throughout the task. In other words, *w* was nor more precise nor more informative than general accuracy. The advantages of using this parameter are thus disputable, yet it is commonly used and referred to in the literature as an appropriate tool to compare data sets from different published studies (e.g., in Castronovo and Göbel, 2012; Halberda et al., 2012; Geary et al., 2015; Libertus et al., 2016).

In the current meta-analysis, we aim at verifying whether Weber fractions computed from various numerical comparison tasks are stable and consistent in the literature. If this were the case, then its usage should be preferred to compare datasets from different studies. Alternatively, the observation of substantial heterogeneity in Weber fractions would be worrying for researchers and for clinicians who want to compare performance from a particular sample or from an individual to some *typical* performance.

## METHODS
### Article Search and Inclusion Criteria
The current meta-analysis only included peer-reviewed articles written in English and published before January 1st 2018 in any scientific journal. Following these inclusion criteria, we independently searched in the three databases *PsycINFO*, *PubMed*, and *Web of Science* for the documents that included the whole expression "*Approximate Number System*" in their title, abstract, keyword, or main body. The cross-referencing of the three searches yielded 387 unique references. We refined the search by looking within each document for any mention of the terms "*Weber*" or "*fraction*". We gathered all matching articles and select the ones that (a) described at least one empirical study conducted on humans with no history of atypical development and that (b) explicitly reported the mean value of the Weber fraction (computed from any non-symbolic comparison task) of their sample(s). Sixty-eight publications were thereby included

in the current meta-analysis. They are further referenced in our bibliography with an asterisk. All statistical analyses were conducted on R Studio (R Core Team R., 2016).

We considered two substantial aspects affecting the evaluation of *w* in the current meta-analysis, in order to minimize any potential risk of bias. First, we highlighted each reported *w* as a function of the mean age of the participants. As noted by Halberda et al. (2012), performance–and subsequently *w*–is intrinsically more heterogeneous in children than in adults (see also, Siegler, 2007). It is consequently insufficient to investigate the variability of this measure within young children. For this reason, we decided to focus on young adults to get a clearer picture of the stability of *w* throughout the literature. Such picture is actually critical to support any claim that Weber fractions are reliable and invariable measures of ANS acuity.

Secondly, and more critically for the purpose of the current meta-analysis, the procedure used to generate stimuli–and to control for non-numerical visual cues–does not have a negligible impact on the value of the Weber fraction (Inglis and Gilmore, 2014; Clayton et al., 2015; Smets et al., 2016). In numerical comparison tasks, participants are sensitive to non-numerical dimensions, and they might base their judgments on them (see Gebuis et al., 2016), so that any systematic confound between the number and one visual property substantially affects behavior. In other words, participants might strategically use available visual information to help them to respond to the task (e.g., the larger array is likely to have more elements). Therefore, paradigms that control for various non-numerical cues at the same time lead to worse performance–and thus larger *w*–than methods involving the manipulation of only one dimension (Smets et al., 2015). The values of *w* reported in a given publication are thus not independent of the properties relative to the methodology used to acquire the data (see for instance, Dietrich et al., 2015b). Although we did not aim for the evaluation of specific influence of a given generation algorithm on participants' performance, we decided to emphasized the properties of the task–and their stimuli–that underlay every considered *w*. However, it should be noted that we did not consider any other methodological aspects that may affect performance (such as the duration of stimulus presentation or the range of the displayed numerical quantities, see Clayton et al., 2015; Smets et al., 2016), as they drastically fluctuated from studies to studies and were thus difficult to categorize in such meta-analysis. We describe how we categorize the dataset in the following section.

## Algorithm Description and Categorization

It is worth noting that a sizeable amount of the retained publications described data collected from more than one participant sample (e.g., comparing different age groups or different methodologies, having different data points in a longitudinal setup). For this reason, we decided to consider data at the sample level, and not at the study level. We then arranged all samples by three categories.

The first category contains the typically developing human samples from publications that explicitly mention the use of the *Panamath*, which is an assessment software freely available

at www.panamath.org. *Panamath* is actually the only existing program that can be implemented with the greatest of ease to test participants or patients, and to directly obtain a performance index, as well as the computation of their Weber fraction. It is thus well known among researchers and practitioners interested in evaluating non-symbolic number abilities. Experimental paradigms of all the samples within this category thus share strong similarities due to the use of the same software. This especially includes the display of two arrays of dots with different colors (blue and yellow) at the same time (see Halberda et al., 2008; for further methodological details). Nonetheless, there may still be some dissimilarity between the experimental conditions because *Panamath* allows researchers to modify some stimuli properties at their best convenience, such as the display duration and the maximal array size. It should be noted that these adaptations are primarily intended to account for the potential youth of the subject taking the test. Anyway, we disregarded such slight modifications in the current meta-analysis, and we considered all samples assessed with *Panamath* in one category.

The second category comprises the samples from studies following Dehaene's et al. (unpublished manuscript) recommendations to construct their stimuli (from Piazza et al., 2004). The authors highlighted in their manuscript that some visual properties are inherently confounded with the number of items in an array. For instance, the picture with the largest number of items is expected to occupy the largest area and/or to possess on average the smallest elements. For this reason, they suggested using a generation algorithm designed to maintain constant one visual property across both displayed arrays, so that this dimension could not be informative to make the decision. Typically, such scripts either consider the individual item size (IIS) or the total occupied area (TOA). However, because number (*N*) is the multiplicative factor between these two parameters, such as IIS × N = TOA, the dimension that is not kept constant across the arrays is systematically correlated with number. To overcome this limitation, Dehaene et al. (unpublished manuscript) recommended generating exactly half of the stimuli with one constant dimension, and the other half with the second unvarying parameter, so that a participant that would strategically make use of the information from one non-numerical parameter would obtain the correct answer in only 50% of the case (which is the chance level). In our meta-analysis, we labeled these programs as "One-dimensional" algorithm as they control for one visual dimension at a time. Noticeably, the *Panamath* software follows this creation rule, as it controls for one visual dimension at a time. Yet the item sizes within one given array are not constant in *Panamath*. We thus excluded this script from the second category and we only considered here studies following the half IIS/half TOA constant rule from Dehaene et al. (unpublished manuscript), without any further restriction.

In the third and last category, we considered all other studies that did not use any of the previously described generation scripts. It is noteworthy that none of these manuscripts put aside the methodological concern that many visual dimensions are inherently confounded with number. On the contrary, their generation procedures all featured their own consideration for

controlling for more than one visual parameter at a time (besides IIS and TOA, such as the length of the convex hull formed by the array or the item density). Among these procedures, one could refer to Gebuis and Reynvoet's (2011) program that manipulates the congruity (or incongruity) of five different dimensions with number throughout the stimuli, to the paradigm of Mussolin et al. (2012) that used collections with various elements richer than single dots, and to the paradigm of DeWind et al. (2015) that disentangles the relative contribution of three orthogonal dimensions (number, spacing, and size) within participants' performance. As these methods accounted for more than one-dimension at a time, we labeled these "Multi-dimensional" algorithms.

## RESULTS

## Description of the Considered Samples

Within the 68 scientific publications that were considered in the current meta-analysis, we retained 115 samples of typically developing humans. Nineteen documents that together described data from 28 samples explicitly mentioned using the *Panamath*. Thirty-six articles were included in the "One-dimensional" category, for a total of 63 typical samples that used such generation algorithm. The third category contained 15 documents that reported on 24 typical samples that used "Multi-dimensional" programs. Descriptive data of considered documents and samples are indicated in **Table 1**. The whole list is

**TABLE 1 |** Description of the data, as a function of the generation algorithm.

| Generation algorithm | Count of documents | Count of samples | Count of participants | Mean sample age (*SD*) | Mean weber fraction (*SD*) |
|---|---|---|---|---|---|
| Panamath | 19 | 28 | 1975 | 12.9 years (*11.7*) | 0.29 (*0.17*) |
| One-dimensional | 36 | 63 | 5230 | 15.1 years (*10.9*) | 0.30 (*0.18*) |
| Multi-dimensional | 15 | 24 | 882 | 14.8 years (*9.21*) | 0.33 (*0.16*) |
| Total | 70[a] | 115 | 8087 | 14.5 years (*10.7*) | 0.30 (*0.17*) |

[a]*One manuscript (Smets et al., 2016) contained two studies that used two different generation algorithms, one one-dimensional and one multi-dimensional. Another document (Smets et al., 2015) reported on two one-dimensional conditions and one multi-dimensional condition. They are thus considered twice in the count of documents in Table 1. Values in italic (and in brackets) are Standard Deviations, which is specified in the column title by their common acronym as (SD).*
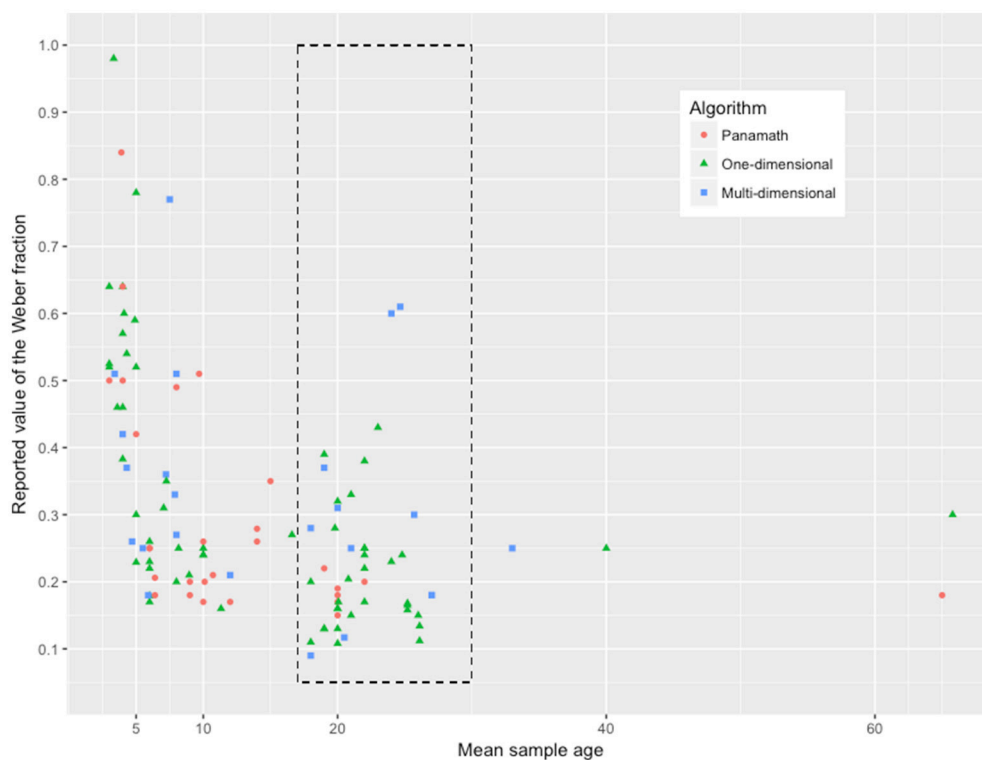


**FIGURE 1 |** Values of the Weber fraction (from 115 typical samples) as a function of mean sample age. We here distinguish Weber fractions depending on the algorithm that was used to measure them (red dot: Panamath; green triangle: One-dimensional algorithm; blue square: Multi-dimensional program). The dashed rectangle encompasses the values from typical adult participants (aged from 18 to 30 years old), which we further consider in **Figure 2**.

available in **Supplementary Table S1**. Overall, mean sample age was 14.5 years, 95% *CI* [12.5, 16.5], and mean Weber fraction value was 0.30, 95% *CI* [0.27, 0.34]. Two one-way analyses of variance–with Generation algorithm as group factor–revealed that both mean sample age and mean Weber fraction did not significantly differ between the three generation algorithms, both $Fs_{(2, 112)} < 1$. We finally conducted an ANCOVA on the value of *w* controlling for mean age, with Generation algorithm as the group factor; this analysis did not lead to any significant effect, $F_{(3, 111)} < 1$.
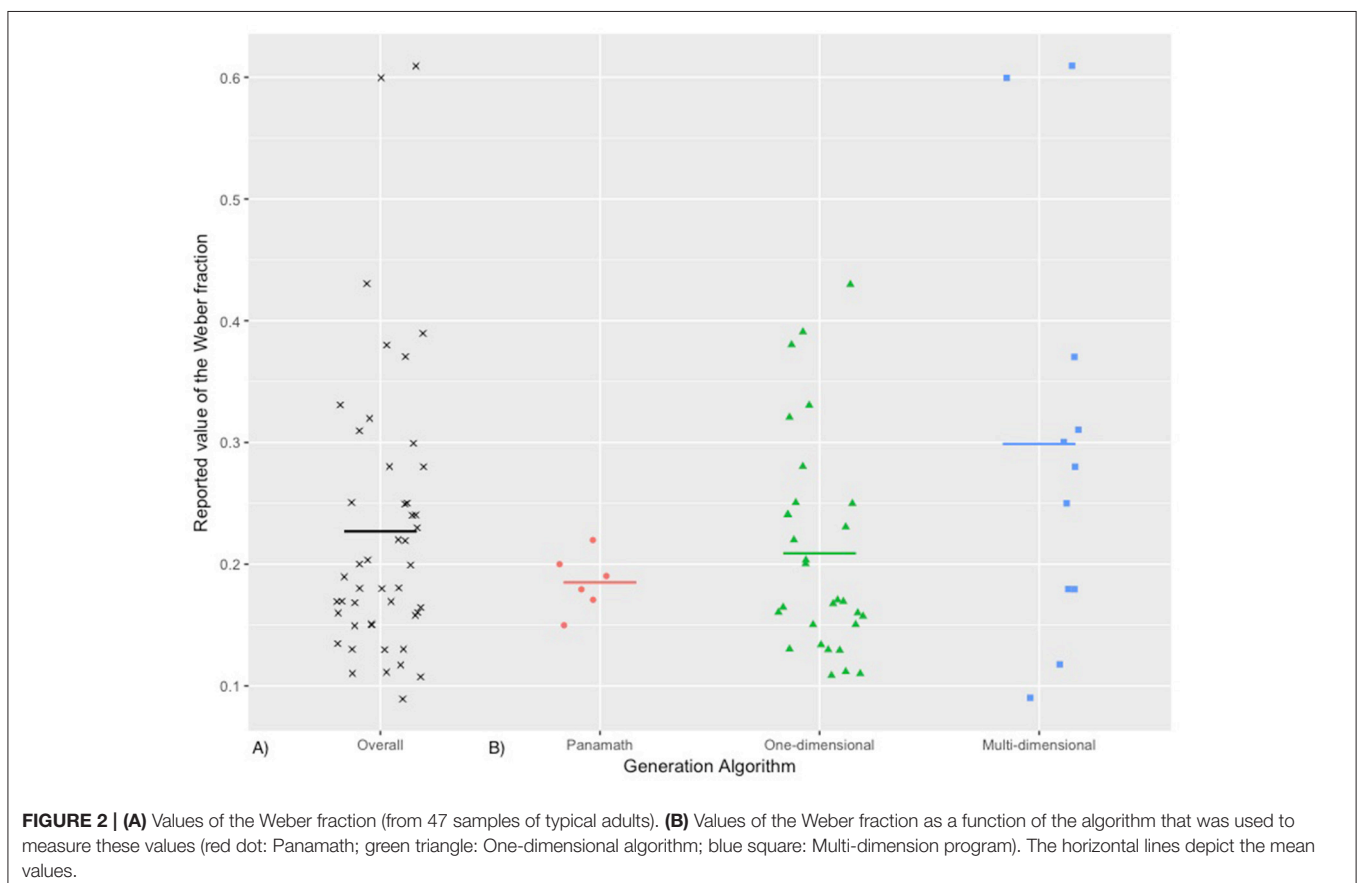
## Weber Fractions in Adults

Despite the overall absence of a significant difference between the three algorithm categories in terms of Weber fractions, a closer look at **Figure 1** revealed that *w* means were not totally independent from mean sample age. Pearson correlation coefficient between the two variables was at $r = -0.41$, which was indeed significant, $t_{(113)} = -4.662$, $p < 0.001$. The value of the Weber fraction thus diminished when age increased, which was in line with previous findings that the noisier numerical acuity at younger age is going through some developmental changes and gradually refines over the years (until ∼30 years, Halberda et al., 2012). Moreover, due to the inherent variability of data collected in children (Siegler, 2007), we focused our further analyses on samples of adults ranging

from 18 to 30 years, in order to be able to compare similar data.

Mean Weber fractions from these selected samples are depicted in **Figure 2A**. Data was collected from 34 documents comprising 47 typically developed adult samples. Mean sample age was 21.68, 95% *CI* [20.89, 22.46], and mean sample size was 48, 95% *CI* [35, 60]. Critically, the mean Weber fraction was 0.22, 95% *CI* [0.19, 0.26]. The latter value drastically ranged, from a minimal value of 0.09 ("congruent condition" from Smets et al., 2015) to a maximal value of 0.61 (in Dietrich et al., 2016). Even in young adult samples, which are expected to be stable, ANS acuity was thus prone to depict substantial heterogeneity.

Such variability seemed to be relative to the generation rule that was followed to create the stimuli (see **Figure 2B**). The *Panamath* lead to the smallest average value: $w = 0.18$, 95% *CI* [0.15, 0.21]. Studies with any One-dimensional algorithm observed a mean *w* value of 0.20, 95% *CI* [0.17, 0.24]. On the other hand, Multi-dimensional algorithms entailed the largest mean *w* value of 0.29, 95% *CI* [0.18, 0.41]. An unilateral Welch test of equality of means revealed that the algorithm category impacted the mean value of the Weber fraction, $F_{(2, 20.388)} = 2.768$, $p = 0.043$. Pairwise comparison tests (Bonferroni corrected) revealed that the Panamath and the One-dimensional category did not significantly differ from each other, $p = 0.627$; however, Multi-dimensional algorithms were significantly greater than the other two, $p = 0.024$ and $p =$



**FIGURE 2 | (A)** Values of the Weber fraction (from 47 samples of typical adults). **(B)** Values of the Weber fraction as a function of the algorithm that was used to measure these values (red dot: Panamath; green triangle: One-dimensional algorithm; blue square: Multi-dimension program). The horizontal lines depict the mean values.

0.046 respectively for the Panamath and the One-dimensional type. Furthermore, we conducted a Brown-Forsythe test for homogeneity of variance, and this test revealed that variance was statistically different between the algorithms, $F_{(2, 44)} = 3.965$, $p = 0.026$. This confirms that the variability of the values was different between the three categories. In other words, as depicted in **Figure 2B**, *Panamath* was less variable than the other generation scripts.

Incidentally, in the publications considered in previous analysis, there were 11 explicit reports of Pearson correlation coefficient between the Weber fraction and overall accuracy in the numerical comparison task. Reported coefficient were very high as the mean $r = 0.97$, 95% $CI$ [0.96, 0.98], with a minimum $r$ of 0.90. This is not surprising, as Weber fractions are computed from accuracy scores (Piazza et al., 2004; Halberda and Feigenson, 2008). As Inglis and Gilmore (2014) pointed out, such high correlation coefficients question whether the Weber fraction is more informative than general accuracy score, and whether the former should be preferred over the latter.

## DISCUSSION

In the current meta-analysis, we highlighted that the Weber fractions computed from numerical comparison tasks are heterogeneous, even within young adult samples. This variability does not support the view that $w$ is a stable parameter devoid of context. As many authors surmised, methodological specificities of the numerical tasks used to compute $w$ impacted its value (e.g., Clayton et al., 2015; Smets et al., 2016), and we were able to characterize this substantial heterogeneity in the literature. As depicted in **Figures 1**, **2**, the method used to generate the non-symbolic arrays substantially affected the mean and the variance of the values of $w$. Multi-dimensional algorithms led to larger $w$ than the other generation programs. This is likely due to the strategic use–or the unconscious experience–of the non-numerical information that is automatically extracted in the visual cortex during the task (Gebuis et al., 2016; Leibovich et al., 2016). At this point, we want to emphasize that we did not aim for the exhaustive description of all methodological discrepancies that might affect measures of ANS acuity (see Dietrich et al., 2015b). For instance, we did not analyse the impact of the range of the quantities used in each study in their evaluation of $w$. In addition, the current meta-analysis did not provide any theoretical evidence that the ANS does not exist (Leibovich et al., 2016; see alternative view from Gebuis et al., 2016). Our analysis only provides evidence that $w$ is not an invariable measure, which may explain some substantial parts in the relation (or non-relation) between ANS acuity and math ability (e.g., Price et al., 2012; Gilmore et al., 2013; Sasanguie et al., 2013).

Interestingly, studies that specifically used the *Panamath* reported homogeneous results. This suggests that the slight methodological difference that we did not consider between these studies–such as the stimulus duration or the numerical range of the arrays–did not drastically impact the measurement of $w$. In other words, *Panamath* studies were thus robust to small dissimilarities in evaluating $w$. It should be noted that our analysis does not simply imply that the *Panamath* reliably assess ANS acuity (see Gebuis et al., 2016; for more detailed methodological considerations). Some might claim that studies from the same laboratory or that use the same exact paradigm arguably tend to show overall higher consistencies, independently from the nature of the task. That being said, our meta-analysis supports that it is possible to reliably measure the same cognitive process in similar numerical comparison tasks, which is reassuring for the literature. It is indeed essential to ascertain that different studies are assessing the same cognitive process before drawing further conclusions about ANS acuity and math ability (see Maxwell et al., 2015; for further considerations about the relevance of replication studies).

Finally, in line with Inglis and Gilmore, 2014 observation, $w$ strongly correlated with general accuracy in the literature. It is unsurprising, as $w$ indexes *in fine* participant accuracy throughout a numerical task. Yet accuracy is modulated by the way non-numerical visual cues are manipulated, with lower performance when multiple visual dimensions are manipulated at the same time (Smets et al., 2015, 2016). The Weber fraction thus does not provide any additional information about performance than overall accuracy does, mostly when taking a correlational perspective. As Inglis and Gilmore (2014) emphasized, one may wonder whether we should compute $w$ at all in the future. With the exception of precise psychophysiological modeling of datasets to highlight specific contribution of numerical and non-numerical dimensions on human behavior (as in DeWind et al., 2015), we believe that most researchers and most clinicians should not bother computing $w$. On the contrary, emphasizing $w$ might give the false impression of its invariability, which might incorrectly encourage direct comparison of very different datasets, whereas reporting percentages of correct responses would not favor such direct comparison. This is not trivial, as the evaluation and the training of ANS acuity both have a substantial clinical impact in the assessment and the remediation of math disability (Mazzocco et al., 2011; Park and Brannon, 2013). We are thus in disfavor of the systematic use of the Weber fraction and in favor of the consideration of normative accuracy datasets acquired from the exact same numerical comparison task.

In conclusion, the Weber fraction is an appealing measure of numerical discrimination due to its psychophysiological nature. It is a precious tool to precisely model human behavior. However, researchers and clinicians should not be unaware of its heterogeneity and its context-depend essence. The algorithm used to generate the stimulus set within the task substantially affects its value and its variability. This measure is thus not directly transferable from one study to another. Researchers and practitioners should thus be extremely cautious when comparing comparison tasks.

## AUTHOR CONTRIBUTIONS

MG and AV: original idea and revision; MG: data collection, data analysis, and drafting manuscript.

## FUNDING

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01694/full#supplementary-material

## REFERENCES

*Anobile, G., Cicchini, G. M., and Burr, D. C. (2014). Separate mechanisms for perception of numerosity and density. *Psychol. Sci.* 25, 265–270. doi: 10.1177/0956797613501520

*Attout, L., Noël, M.-P., Vossius, L., and Rousselle, L. (2017). Evidence of the impact of visuo-spatial processing on magnitude representation in 22q11.2 microdeletion syndrome. *Neuropsychologia* 99, 296-305. doi: 10.1016/j.neuropsychologia.2017.03.023

Barth, H., La Mont, K., Lipton, J. S., Dehaene, S., Kanwisher, N., and Spelke, E. S. (2006). Non-symbolic arithmetic in adults and young children. *Cognition* 98, 199-222. doi: 10.1016/j.cognition.2004.09.011

*Bugden, S., and Ansari, D. (2016). Probing the nature of deficits in the "Approximate Number System" in children with persistent Developmental Dyscalculia. *Dev. Sci.* 19, 817–833. doi: 10.1111/desc.12324

Cantlon, J. F., Cordes, S., Libertus, M. E., and Brannon, E. M. (2009). Comment on "Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures". *Science* 323:38b. doi: 10.1126/science.1156540

*Cappelletti, M., Didino, D., Stoianov, I., and Zorzi, M. (2014). Number skills are maintained in healthy ageing. *Cogn. Psychol.* 69, 25–45. doi: 10.1016/j.cogpsych.2013.11.004

Castronovo, J., and Göbel, S. M. (2012). Impact of high mathematics education on the number sense. PLoS ONE 7:e33832. doi: 10.1371/journal.pone.0033832

*Chesney, D., Bjalkebring, P., and Peters, E. (2015). How to estimate how well people estimate: evaluating measures of individual differences in the approximate number system. *Attention Percept. Psychophys.* 77, 2781–2802. doi: 10.3758/s13414-015-0974-6

*Chu, F. W., vanMarle, K., and Geary, D. C. (2013). Quantitative deficits of preschool children at risk for mathematical learning disability. *Front. Psychol.* 4:195. doi: 10.3389/fpsyg.2013.00195

Cicchini, G. M., Anobile, G., and Burr, D. C. (2014). Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proc. Natl. Acad. Sci.* 111, 7867-7872. doi: 10.1073/pnas.1402785111

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

Clayton, S., Inglis, M., and Gilmore, C. (2018). Developmental differences in approaches to nonsymbolic comparison tasks. *Quart. J. Exp. Psychol.* 1:1747021818755296. doi: 10.1177/1747021818755296

Dehaene, S. (1997). *The Number Sense: How the Mind creates Mathematic.* New York, NY: Oxford University Press.

Dehaene, S. (2003). The neural basis of the Weber–Fechner law: a logarithmic mental number line. *Trends Cogn. Sci.* 7, 145-147. doi: 10.1016/S1364-6613(03)00055-X

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

*DeWind, N. K., and Brannon, E. M. (2016). Significant inter-test reliability across approximate number system assessments. *Front. Psychol.* 7:310. doi: 10.3389/fpsyg.2016.00310

*Dietrich, J. F., Huber, S., Klein, E., Willmes, K., Pixner, S., and Moeller, K. (2016). A systematic investigation of accuracy and response time based measures used to index ANS acuity. *PLoS ONE* 11:e0163076. doi: 10.1371/journal.pone.0163076

*Dietrich, J. F., Huber, S., Moeller, K., and Klein, E. (2015a). The influence of math anxiety on symbolic and non-symbolic magnitude processing. *Front. Psychol.* 6:1621. doi: 10.3389/fpsyg.2015.01621

Dietrich, J. F., Huber, S., and Nuerk, H.-C. (2015b). Methodological aspects to be considered when measuring the approximate number system (ANS): a research review. *Front. Psychol.* 6:295. doi: 10.3389/fpsyg.2015.00295

*Fazio, L. K., Bailey, D. H., Thompson, C. A., and Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *J. Exp. Child Psychol.* 123, 53-72. doi: 10.1016/j.jecp.2014.01.013

Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core systems of number. *Trends Cogn. Sci.* 8, 307-314. doi: 10.1016/j.tics.2004.05.002

Gallistel, C. R., and Gelman, R. (2000). Non-verbal numerical cognition: from reals to integers. *Trends Cogn. Sci.* 4, 59–65. doi: 10.1016/S1364-6613(99)01424-2

*Geary, D. C., Hoard, M. K., Nugent, L., and Rouder, J. N. (2015). Individual differences in algebraic cognition: relation to the approximate number and semantic memory systems. *J. Exp. Child Psychol.* 140, 211–227. doi: 10.1016/j.jecp.2015.07.010

Gebuis, T., Cohen Kadosh, R., and Gevers, W. (2016). Sensory-integration system rather than approximate number system underlies numerosity processing: a critical review. *Acta Psychol.* 171, 17–35. doi: 10.1016/j.actpsy.2016.09.003

Gebuis, T., and Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behav. Res. Methods* 43, 981–986. doi: 10.3758/s13428-011-0097-5

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., et al. Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS ONE* 8:e67374. doi: 10.1371/journal.pone.0067374

*Gilmore, C., Attridge, N., and Inglis, M. (2011). Measuring the approximate number system. *Q. J. Exp. Psychol.* 64, 2099–2109. doi: 10.1080/17470218.2011.574710

*Gomez, A., Piazza, M., Jobert, A., Dehaene-Lambertz, G., Dehaene, S., and Huron, C. (2015). Mathematical difficulties in developmental coordination disorder: symbolic and nonsymbolic number processing. *Res. Dev. Disabilities* 43–44, 167–178. doi: 10.1016/j.ridd.2015.06.011

*Guillaume, M., Gevers, W., and Content, A. (2016). Assessing the approximate number system: no relation between numerical comparison and estimation tasks. *Psychol. Res.* 80, 248–258. doi: 10.1007/s00426-015-0657-x

*Guillaume, M., Nys, J., Mussolin, C., and Content, A. (2013). Differences in the acuity of the Approximate Number System in adults: the effect of mathematical ability. *Acta Psychol.* 144, 506–512. doi: 10.1016/j.actpsy.2013.09.001

*Halberda, J., and 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. *Dev. Psychol.* 44, 1457–1465. doi: 10.1037/a0012682

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., and Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11116-11120. doi: 10.1073/pnas.1200196109

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

*Hellgren, K., Halberda, J., Forsman, L., Ådén, U., and Libertus, M. (2013). Compromised approximate number system acuity in school-aged children born extremely preterm. *Dev. Med. Child Neurol.* 55, 1070–1071. doi: 10.1111/dmcn.12206

*Inglis, M., and Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychol.* 145, 147–155. doi: 10.1016/j.actpsy.2013.11.009

*Jang, S., and Cho, S. (2016). The acuity for numerosity (but not continuous magnitude) discrimination correlates with quantitative problem solving but not routinized arithmetic. *Curr. Psychol.* 35, 44–56. doi: 10.1007/s12144-015-9354-6

*Jones, S. M., Pearson, J., DeWind, N. K., Paulsen, D., Tenekedjieva, A.-M., and Brannon, E. M. (2014). Lemurs and macaques show similar numerical sensitivity. *Anim. Cogn.* 17, 503–515. doi: 10.1007/s10071-013-0682-3

Julio-Costa, A., Starling-Alves, I., Lopes-Silva, J. B., Wood, G., and Haase, V. G. (2015). Stable measures of number sense accuracy in math learning disability: Is it time to proceed from basic science to clinical application? *Psych J.* 4, 218–225. doi: 10.1002/pchj.114

*Khanum, S., Hanif, R., Spelke, E. S., Berteletti, I., and Hyde, D. C. (2016). Effects of non-symbolic approximate number practice on symbolic numerical abilities in pakistani children. *PLoS ONE* 11:e0164436. doi: 10.1371/journal.pone.0164436

*Knops, A., Dehaene, S., Berteletti, I., and Zorzi, M. (2014). Can approximate mental calculation account for operational momentum in addition and subtraction? *Quart. J. Exp. Psychol.* 67, 1541–1556. doi: 10.1080/17470218.2014.890234

Leibovich, T., Katzin, N., Harel, M., and Henik, A. (2016). From "sense of number" to "sense of magnitude": The role of continuous magnitudes in numerical cognition. *Behav. Brain Sci.* 40, 1–62. doi: 10.1017/S0140525X16000960

*Libertus, M. E., Feigenson, L., and Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability: approximate number system and math abilities. *Dev. Sci.* 14, 1292–1300. doi: 10.1111/j.1467-7687.2011.01080.x

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

*Libertus, M. E., Feigenson, L., Halberda, J., and Landau, B. (2014). Understanding the mapping between numerical approximation and number words: evidence from Williams syndrome and typical development. *Dev. Sci.* 17, 905–919. doi: 10.1111/desc.12154

*Libertus, M. E., Odic, D., Feigenson, L., and Halberda, J. (2016). The precision of mapping between number words and the approximate number system predicts children's formal math abilities. *J. Exp. Child Psychol.*, 150, 207–226. doi: 10.1016/j.jecp.2016.06.003

*Libertus, M. E., Odic, D., and Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychol.* 141, 373–379. doi: 10.1016/j.actpsy.2012.09.009

*Lindskog, M., Winman, A., and Juslin, P. (2014). The association between higher education and approximate number system acuity. *Front. Psychol.* 5:462. doi: 10.3389/fpsyg.2014.00462

*Lindskog, M., Winman, A., Juslin, P., and Poom, L. (2013). Measuring acuity of the approximate number system reliably and validly: the evaluation of an adaptive test procedure. *Front. Psychol.* 4:510. doi: 10.3389/fpsyg.2013.00510

*Lindskog, M., Winman, A., and Poom, L. (2016). Arithmetic Training does not improve approximate number system acuity. *Front. Psychol.* 7:1634. doi: 10.3389/fpsyg.2016.01634

*Liu, Y. (2017). Fraction magnitude understanding and its unique role in predicting general mathematics achievement at two early stages of fraction instruction. *Br. J. Educ. Psychol.* 88:345–362. doi: 10.1111/bjep.12182

*Lonnemann, J., Linkersdörfer, J., Hasselhorn, M., and Lindberg, S. (2013). Developmental changes in the association between approximate number representations and addition skills in elementary school children. *Front. Psychol.* 4:783. doi: 10.3389/fpsyg.2013.00783

*Lourenco, S. F., Bonny, J. W., Fernandez, E. P., and Rao, S. (2012). Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence. *Proc. Natl. Acad. Sci.* 109, 18737–18742. doi: 10.1073/pnas.1207212109

*Lyons, I. M., and Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition* 121, 256-261. doi: 10.1016/j.cognition.2011.07.009

Maxwell, S. E., Lau, M. Y., and Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *Am. Psychol.* 70, 487-498. doi: 10.1037/a0039400

*Mazzocco, M. M. M., Feigenson, L., and Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (Dyscalculia): impaired numerical acuity contributes to MLD. *Child Dev.* 82, 1224–1237. doi: 10.1111/j.1467-8624.2011.01608.x

*Mussolin, C., Nys, J., Content, A., and Leybaert, J. (2014). Symbolic number abilities predict later approximate number system acuity in preschool children. *PLoS ONE* 9:e91839. doi: 10.1371/journal.pone.0091839

*Mussolin, C., Nys, J., Leybaert, J., and Content, A. (2012). Relationships between approximate number system acuity and early symbolic number abilities. *Trends Neurosci. Educ.* 1, 21-31. doi: 10.1016/j.tine.2012.09.003

*Norris, J. E., and Castronovo, J. (2016). Dot Display Affects approximate number system acuity and relationships with mathematical achievement and inhibitory control. *PLoS ONE* 11:e0155543. doi: 10.1371/journal.pone.0155543

*Norris, J. E., McGeown, W. J., Guerrini, C., and Castronovo, J. (2015). Aging and the number sense: preserved basic non-symbolic numerical processing and enhanced basic symbolic processing. *Front. Psychol.* 6:999. doi: 10.3389/fpsyg.2015.00999

*Núñez, R. E. (2017). Is there really an evolved capacity for number? *Trends Cogn. Sci.* 21, 409-424. doi: 10.1016/j.tics.2017.03.005

*Odic, D. (2017). Children's intuitive sense of number develops independently of their perception of area, density, length, and time. *Dev. Sci.* 21:e12533. doi: 10.1111/desc.12533

*Odic, D., and Halberda, J. (2015). Eye movements reveal distinct encoding patterns for number and cumulative surface area in random dot arrays. *J. Vis.* 15:5. doi: 10.1167/15.15.5

*Odic, D., Hock, H., and Halberda, J. (2014). Hysteresis affects approximate number discrimination in young children. *J. Exp. Psychol.* 143, 255–265. doi: 10.1037/a0030825

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

*Odic, D., Pietroski, P., Hunter, T., Lidz, J., and Halberda, J. (2013b). Young children's understanding of ?more? and discrimination of number and surface area. *J. Exp. Psychol.* 39, 451–461. doi: 10.1037/a0028874

*Oliveira, L. F. S., Santos, A. O., Vianna, G. S., Di Ninno, C. Q. M. S., Giacheti, C. M., Carvalho, M. R. S., et al. Haase, V. G. (2014). Impaired acuity of the approximate number system in 22q11.2 microdeletion syndrome. *Psychol. Neurosci.* 7, 151–158. doi: 10.3922/j.psns.2014.02.04

*Olsson, L., Östergren, R., and Träff, U. (2016). Developmental dyscalculia: a deficit in the approximate number system or an access deficit? *Cogn. Dev.*, 39, 154–167. doi: 10.1016/j.cogdev.2016.04.006

Park, J., and Brannon, E. M. (2013). Training the Approximate number system improves math proficiency. *Psychol. Sci.* 24, 2013-2019. doi: 10.1177/0956797613482944

*Park, J., and Starns, J. J. (2015). The approximate number system acuity redefined: a diffusion model approach. *Front. Psychol.* 6:1955. doi: 10.3389/fpsyg.2015.01955

*Patalano, A. L., Saltiel, J. R., Machlin, L., and Barth, H. (2015). The role of numeracy and approximate number system acuity in predicting value and probability distortion. *Psychon. Bull. Rev.* 22, 1820–1829. doi: 10.3758/s13423-015-0849-9

*Piantadosi, S. T. (2016). A rational analysis of the approximate number system. *Psychon. Bull. Rev.* 23, 877–886. doi: 10.3758/s13423-015-0963-8

*Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., et al. Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition* 116, 33–41. doi: 10.1016/j.cognition.2010.03.012

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., and Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron* 44, 547–555. doi: 10.1016/j.neuron.2004.10.014

Piazza, M., Pica, P., Izard, V., Spelke, E. S., and Dehaene, S. (2013). Education enhances the acuity of the nonverbal approximate number system. *Psychol. Sci.* 24, 1037-1043. doi: 10.1177/0956797612464057

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

*Pietroski, P., Lidz, J., Hunter, T., and Halberda, J. (2009). The meaning of "most": Semantics, numerosity and psychology. *Mind Lang.* 24, 554–585. doi: 10.1111/j.1468-0017.2009.01374.x

*Pinheiro-Chagas, P., Wood, G., Knops, A., Krinzinger, H., Lonnemann, J., Starling-Alves, I., et al. (2014). In how many ways is the approximate number system associated with exact calculation? *PLoS ONE* 9:e111155. doi: 10.1371/journal.pone.0111155

Platt, J. R., and Johnson, D. M. (1971). Localization of position within a homogeneous behavior chain: effects of error contingencies. *Learn. Motiv.* 2, 386-414.

*Price, G. R., Palmer, D., Battista, C., and 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 Psychol.* 140, 50–57. doi: 10.1016/j.actpsy.2012.02.008

R Core Team R. (2016). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing,* Vienna. Available online at: https://www.R-project.org

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., and Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number–space mappings: what underlies mathematics achievement? *J. Exp. Child Psychol.* 114, 418-431. doi: 10.1016/j.jecp.2012.10.012

*Shusterman, A., Slusser, E., Halberda, J., and Odic, D. (2016). Acquisition of the cardinal principle coincides with improvement in approximate number system acuity in preschoolers. *PLoS ONE* 11:e0153072. doi: 10.1371/journal.pone.0153072

Siegler, R. S. (2007). Cognitive variability. *Dev. Sci.*, 10, 104-109. doi: 10.1111/j.1467-7687.2007.00571.x

*Skagerlund, K., and Träff, U. (2014). Development of magnitude processing in children with developmental dyscalculia: space, time, and number. *Front. Psychol.* 5:675. doi: 10.3389/fpsyg.2014.00675

*Skagerlund, K., and Träff, U. (2016). Processing of space, time, and number contributes to mathematical abilities above and beyond domain-general cognitive abilities. *J. Exp. Child Psychol.* 143, 85-101. doi: 10.1016/j.jecp.2015.10.016

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

*Smets, K., Moors, P., and Reynvoet, B. (2016). Effects of presentation type and visual control in numerosity discrimination: implications for number processing? *Front. Psychol.* 7:66. doi: 10.3389/fpsyg.2016.00066

*Smets, K., Sasanguie, D., Szücs, D., and Reynvoet, B. (2015). The effect of different methods to construct non-symbolic stimuli in numerosity estimation and comparison. *J. Cogn. Psychol.* 27, 310–325. doi: 10.1080/20445911.2014.996568

*Starr, A., Libertus, M. E., and Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proc. Natl. Acad. Sci.* 110, 18116-18120. doi: 10.1073/pnas.1302751110

Stevens, S. S. (1957). On the psychophysical law. *Psychol. Rev.* 64:153. doi: 10.1037/h0046162

*Szucs, D., Nobes, A., Devine, A., Gabriel, F. C., and Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. *Front.Psychol.* 4:444. doi: 10.3389/fpsyg.2013.00444

*Tinelli, F., Anobile, G., Gori, M., Aagten-Murphy, D., Bartoli, M., Burr, D. C., et al. Concetta Morrone, M. (2015). Time, number and attention in very low birth weight children. *Neuropsychologia* 73, 60–69. doi: 10.1016/j.neuropsychologia.2015.04.016

*Tosto, M. G., Petrill, S. A., Halberda, J., Trzaskowski, M., Tikhomirova, T. N., Bogdanova, O. Y., et al. (2014). Why do we differ in number sense? Evidence from a genetically sensitive investigation. *Intelligence* 43, 35-46. doi: 10.1016/j.intell.2013.12.007

*van den Berg, R., Lindskog, M., Poom, L., and Winman, A. (2017). Recent is more: a negative time-order effect in nonsymbolic numerical judgment. *J. Exp. Psychol.* 43, 1084–1097. doi: 10.1037/xhp0000387

Van Oeffelen, M. P., and Vos, P. G. (1982). A probabilistic model for the discrimination of visual number. *Percept. Psychophys.* 32, 163–170. doi: 10.3758/BF03204275

*vanMarle, K., Chu, F. W., Li, Y., and Geary, D. C. (2014). Acuity of the approximate number system and preschoolers? Quantitative development. *Dev. Sci.* 17, 492–505. doi: 10.1111/desc.12143

*Whalen, J., Gallistel, C. R., and Gelman, R. (1999). Nonverbal counting in humans: the psychophysics of number representation. *Psychol. Sci.* 10, 130-137.

*Winman, A., Juslin, P., Lindskog, M., Nilsson, H., and Kerimi, N. (2014). The role of ANS acuity and numeracy for the calibration and the coherence of subjective probability judgments. *Front. Psychol.* 5:851. doi: 10.3389/fpsyg.2014.00851

*Wong, T. T., Ho, C. S.-H., and Tang, J. (2014). Identification of children with mathematics learning disabilities (MLDs) using latent class growth analysis. *Res. Dev. Disabil.* 35, 2906–2920. doi: 10.1016/j.ridd.2014.07.015

*Wong, T. T.-Y., Ho, C. S.-H., and Tang, J. (2016). The relation between ANS and symbolic arithmetic skills: the mediating role of number-numerosity mappings. *Contemp. Educ. Psychol.* 46, 208–217. doi: 10.1016/j.cedpsych.2016.06.003

---

*Publications included in the meta-analysis are referenced with an asterisk.