Developmental Science

Developmental Science (2016), pp 1–16 DOI: 10.1111/desc.12372

PAPER

Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis

Michael Schneider,[1] Kassandra Beeres,[1] Leyla Coban,[1] Simon Merz,[1] S. Susan Schmidt,[1] Johannes Stricker[1] and Bert De Smedt[2]

1. *Department of Psychology, University of Trier, Germany*
2. *Faculty of Psychology and Educational Sciences, Parenting and Special Education Research Group, KU Leuven, University of Leuven, Belgium*

Abstract

Many studies have investigated the association between numerical magnitude processing skills, as assessed by the numerical magnitude comparison task, and broader mathematical competence, e.g. counting, arithmetic, or algebra. Most correlations were positive but varied considerably in their strengths. It remains unclear whether and to what extent the strength of these associations differs systematically between non-symbolic and symbolic magnitude comparison tasks and whether age, magnitude comparison measures or mathematical competence measures are additional moderators. We investigated these questions by means of a meta-analysis. The literature search yielded 45 articles reporting 284 effect sizes found with 17,201 participants. Effect sizes were combined by means of a two-level random-effects regression model. The effect size was significantly higher for the symbolic (r = .302, 95% CI [.243, .361]) than for the non-symbolic (r = .241, 95% CI [.198, .284]) magnitude comparison task and decreased very slightly with age. The correlation was higher for solution rates and Weber fractions than for alternative measures of comparison proficiency. It was higher for mathematical competencies that rely more heavily on the processing of magnitudes (i.e. mental arithmetic and early mathematical abilities) than for others. The results support the view that magnitude processing is reliably associated with mathematical competence over the lifespan in a wide range of tasks, measures and mathematical subdomains. The association is stronger for symbolic than for non-symbolic numerical magnitude processing. So symbolic magnitude processing might be a more eligible candidate to be targeted by diagnostic screening instruments and interventions for school-aged children and for adults.

Research highlights

- This is the first meta-analysis on the association of non-symbolic and symbolic magnitude comparison with mathematical competence.
- The meta-analysis synthesized 284 effect sizes from 17,201 participants by means of a random-effects two-level regression model.
- Associations with mathematical competence were stronger for symbolic than for non-symbolic measures.
- Measures of comparison and mathematical competence strongly moderated the comparison–competence association.

Introduction

A wealth of empirical studies investigated the association between the processing of numerical magnitudes and broader mathematical competence. These studies have far-reaching implications because numbers are of fundamental importance in our society. For example, numerical skills are strong predictors of success in school (Duncan, Dowsett, Claessens, Magnuson, Huston *et al*., 2007), of medical decision making (Reyna, Nelson, Han & Dieckmann, 2009), and of valuations of monetary amounts (Schley & Peters, 2014). They are associated with socioeconomic status (Ritchie & Bates, 2013) and mortgage default (Gerardi, Goette & Meier, 2013).

Address for correspondence: Michael Schneider, University of Trier – Division I, 54286 Trier, Germany; e-mail: m.schneider@uni-trier.de

While previous studies have converged on the conclusion that numerical magnitude processing is an important foundation for higher-level mathematical competence, studies are heterogeneous in respect of whether the processing of non-symbolic magnitude representations (i.e. dots), symbolic magnitude representations (i.e. digits), or both are relevant for the learning of more advanced mathematical competence. Moreover, it has been suggested that the association between magnitude processing and broader mathematical competence might also be moderated by participant age (e.g. Inglis, Attridge, Batchelor & Gilmore, 2011; Rousselle & Noël, 2008), by measures of magnitude comparison skills (e.g. Price, Palmer, Battista & Ansari, 2012), or by measures of mathematical competence (e.g. De Smedt, Noël, Gilmore & Ansari, 2013). This hampers the integration of empirical findings across studies in narrative reviews of the literature (De Smedt *et al.*, 2013; Feigenson, Libertus & Halberda, 2013).

Meta-analyses quantitatively integrate findings across studies and allow for explicit tests of moderating influences of third variables. Therefore, meta-analyses can substantially advance our understanding of the associations between numerical magnitude processing and broader mathematical competence. Two recent meta-analyses have investigated the associations between non-symbolic numerical magnitude processing and mathematical competence. Chen and Li (2014) included 47 effect sizes and found an overall correlation of $r = .20$, 95% CI [.14, .26] for cross-sectional studies. Fazio, Bailey, Thompson and Siegler (2014) included 34 effect sizes and found an overall correlation of $r = .22$, 95% CI [.20, .25].

These two meta-analyses provided strong evidence for a weak but reliable association between non-symbolic magnitude processing and mathematical competence, but they were limited in their scope. They included only one effect size from each sample. So it was not possible to analyze differences between effect sizes found with the same sample, for example, accuracy versus speed of the same participants. Even more crucially, these meta-analyses only included findings obtained with non-symbolic magnitude processing tasks. No conclusions about differences between non-symbolic and symbolic magnitude processing could be drawn. The current meta-analysis closes this gap in the research literature. A better understanding of how non-symbolic and symbolic magnitude processing relate to broader mathematical competence might provide helpful background information for educational interventions aiming at improving learners' numerical processing skills as preparation for more advanced mathematical learning (De Smedt *et al.*, 2013; Feigenson *et al.*, 2013). At a theoretical level, it will aid to evaluate the importance of evolutionary older non-symbolic magnitude representations, which we share with many other species (Cantlon, 2012), compared to uniquely human symbolic representations of numbers.

The current study therefore included effect sizes from non-symbolic and from symbolic numerical magnitude processing tasks. We used a two-level regression model with effect sizes (level 1) nested under independent samples (level 2), so that all effect sizes from all samples could be included. This yielded a database of 284 effect sizes and allowed us to estimate the overall effect sizes along with the moderating influences of a non-symbolic vs. symbolic task format, participant age, comparison measures and mathematical competence measures.

In our meta-analysis we focused on the comparison task, which is the most frequently used task to assess numerical magnitude processing (Ansari, 2008; Dehaene, Dupoux & Mehler, 1990; Moyer & Landauer, 1967). In this task, two dot arrays (non-symbolic) or two numbers presented in the form of Arabic digits (symbolic) are presented and the participant has to indicate the one with the larger (or the smaller) numerical magnitude. The numerosities can be presented simultaneously or sequentially, and either both of them change from trial to trial or one is a fixed standard against which the other numerosity has to be compared.

A narrative review of the literature suggests that the association between numerical magnitude processing and broader mathematical competence might be more robust and consistent for studies with the symbolic magnitude processing tasks than for studies with the non-symbolic task (De Smedt *et al.*, 2013). In that review, 13 out of 17 empirical studies (76%) observed a significant correlation between symbolic magnitude comparison and mathematical competence, whereas only 11 out of 25 studies (44%) found a significant correlation between non-symbolic magnitude comparison and mathematical competence. However, such comparisons are of limited use because they take neither effect size differences nor standard errors nor sample size differences between studies into account, which further highlights the need for including both non-symbolic and symbolic associations in a meta-analysis on these questions.

It remains an open question whether the association between numerical magnitude processing and mathematical competence is moderated by age (Inglis *et al.*, 2011; Rousselle & Noël, 2008). One possibility is that processing of numerical magnitudes as assessed by the comparison task merely provides a starting point for mathematical development so that the correlation between the two constructs decreases with age as environmental influences keep accumulating (Inglis *et al.*, 2011). This is plausible because processing of whole-number magnitudes plays a central role in elementary school mathematics with its focus on arithmetic, but is less important for understanding more advanced mathematical concepts in higher school grades where variables and their abstract interrelations become more important than concrete numerical magnitudes (Schneider, Grabner & Paetsch, 2009). Alternatively, it has been proposed that magnitude processing gives people a number sense which is an integral aspect of nearly all mathematical thinking, so that these constructs should be linked throughout the lifespan (Halberda, Lya, Wilmer, Naiman & Germine, 2012; Libertus, Odic & Halberda, 2012). The two previous meta-analyses investigated the moderating effect of age, yet with inconsistent results. Chen and Li (2014) found no moderating effect of age. However, they only compared children younger than 12 years with adults older than 17 years. Fazio *et al*. (2014) found that participant age moderated the size of the comparison–competence correlation, with the correlation being larger for children younger than 6 (r = .40) than for students between 6 and 18 years (r = .17) and adults (r = .21) who did not differ. For the symbolic magnitude comparison task no meta-analytic results have been published so far.

Another possible moderator of the association between numerical magnitude comparison and mathematical competence is the comparison measure, that is, the operationalization of performance on the comparison task. Typically, proficiency on the comparison task is measured as solution rate or solution time. Solution times are often preferred over solution rates because the latter can yield ceiling effects in older children and adults and thus might lead to an underestimation of the true correlation between comparison and competence (cf. Berch, 2005; Holloway & Ansari, 2009). Alternatively to these general indices, researchers often calculate more specific measures. Performance on the non-symbolic tasks is often characterized by a ratio-effect: The closer the ratio of the compared dot arrays is to 1, the more difficult it is to discriminate the dot arrays. This effect can also be described by the Weber fraction (W), which is the smallest ratio of two numerosities that a person can reliably judge as larger or smaller (Halberda, Mazzocco & Feigenson, 2008). Performance on the symbolic task is often quantified by computing the distance effect, which indicates that accuracy increases and reaction time decreases as the numerical distance (i.e. the difference) between two numbers decreases. This distance effect has been computed either as a standardized difference score for small vs. large numerical distances (e.g. Holloway & Ansari, 2009) or as the slope of a regression in which reaction time is predicted by numerical distance (e.g. De Smedt, Verschaffel & Ghesquière, 2009; Schneider *et al*., 2009). The distance effect was originally interpreted as an index of the overlap between different analogue representations of magnitude (Cohen Kadosh, Henik, Rubinsten, Mohr, Dori *et al*., 2005; Dehaene, 1997; Moyer & Landauer, 1967), yet alternative explanations that point to the role of more general decisional processes have been put forward (Holloway & Ansari, 2008; Van Opstal, Gevers, De Moor & Verguts, 2008). The ratio-effect and distance effect are conceptually similar (Bartelet, Vaessen, Blomert & Ansari, 2014). The inter-correlations of alternative operationalizations of magnitude comparison skills have been found to be much lower than expected in some samples, suggesting that the measures might tap into partly different aspects of comparison skills (Gilmore, Attridge & Inglis, 2011). Relatedly, these measures also differ in their reliabilities (Inglis & Gilmore, 2014; Price *et al*., 2012). It is thus plausible to assume that they also differ in their associations with broader mathematical competence.

A further moderator of the comparison–competence association might be the measure of mathematical competence. Most mathematical competence measures have in common that they quantify proficiency as solution rate. However, mathematical competence measures differ strongly in their content and this might lead to different associations with magnitude processing. For example, it is plausible to assume that numerical magnitude processing skills are important for counting and whole-number arithmetic (Gilmore, McCarthy & Spelke, 2007; Libertus, Feigenson & Halberda, 2011). In contrast, in subdomains such as algebra, numerical magnitudes still play a role, but are arguably less relevant than understanding abstract mathematical concepts, such as variables and equivalence (cf. Schneider, Rittle-Johnson & Star, 2011). It is hard to see how numerical magnitude processing could directly support such non-numerical mathematical concepts. In line with this, algebraic competence correlates higher with an understanding of abstract mathematical concepts than with whole-number processing (Booth & Newton, 2012). We thus expect that the comparison–competence correlation is higher in mathematical subdomains where the mental processing of numerical magnitudes is more important (e.g. mental arithmetic) and lower in other mathematical subdomains, for example, algebra and geometry.

Against this background, the aim of the current meta-analysis was to statistically integrate the available evidence on the association between numerical magnitude processing and broader mathematical competence with a special focus on differences between non-symbolic and symbolic magnitude comparison tasks. In addition to task format, we included age, comparison measure and mathematical competence measure as potential moderators in our analyses and tested to what extent they explained the heterogeneity of the effect sizes reported in the literature.

Method

Literature search and inclusion criteria

We searched the title, abstract, and keywords of all articles in the literature database PsycINFO in October 2014 with the search string ((‘math* achievement’ or ‘math* competence’ or ‘math* skill*’ or ‘math* abilit*’ or ‘math* performance’ or ‘arithmetic*’ or ‘num* skill*’) and (‘magnitude representation*’ or ‘distance effect*’ or ‘approximate number system’ or ‘numerical cognition’ or ‘number sense’ or ‘number acuity’ or ‘digit comparison*’ or ‘number comparison*’)) and limited the results to empirical studies with non-disordered human populations that had been published in a peer-reviewed journal in the English language.

Two trained raters judged independently of each other for each article whether it was included in the meta-analysis. Disagreements were resolved by discussion. The inclusion of studies and of effect sizes within studies in our database was determined by the following criteria: (1) The study reported original empirical findings (i.e. not a re-analysis of already reported findings or a review). (2) The study included at least one numerical magnitude comparison task, which required the participants on each trial to indicate the larger or the smaller of two numerical magnitudes. The numerical magnitudes could be presented sequentially or simultaneously, and as dot arrays or Arabic numerals. The magnitudes could include one-digit and multi-digit integers, but not negative numbers or non-whole numbers. Participants’ behavior on the task had to be coded as solution rate, solution time, distance effect with solution rates, distance effect with solution times, ratio effect with solution rates, ratio effect with solution times, or Weber fraction. (3) The study additionally included at least one measure of mathematical competence other than magnitude comparison, such as the TEMA test of early mathematical ability (Ginsburg & Baroody, 2003), a mental or written arithmetic task, a curriculum-based or a more general standardized test of mathematics, school grades in mathematics, a numerical reasoning test, or a mathematical problem solving task. Measures that are usually interpreted as assessing basic numerical processing (e.g. ordering of magnitudes, same–different judgments, odd–even judgments, naming of numerical magnitudes, number line estimation) were not considered as measures of mathematical competence because they are conceptually too closely related to magnitude comparison. (4) The study reports at least one standardized effect size of the strength and the direction of the bivariate relation between a magnitude comparison measure and a mathematical competence measure. The study also reports the sample size for this effect. Effect sizes from multivariate analyses (multifactorial ANOVAs, multiple regressions, or partial correlations) were not included because their outcomes depend on all variables included in the respective model, which limits the comparability. Authors who exclusively reported multivariate results in a study were asked by email to provide the corresponding Pearson correlations, which were then included in the meta-analysis. (5) The study reported at least one effect size for a sample with a majority of normally developing participants, who had not been diagnosed with dyscalculia or mathematical learning difficulties.

Meta-analyses can be biased by the file-drawer problem, that is, by the fact that statistically significant results have a higher probability of getting published than statistically non-significant results. The authors of some meta-analyses try to solve this problem by searching for and including unpublished studies. We did not do this here, because the quality of unpublished studies is hard to assess and because researchers usually obtain only a non-representative set of unpublished studies (e.g. mostly from their home country or from their own research community) that does not increase the quality of the meta-analytic results (Ferguson & Brannick, 2011). Instead, we accounted for publication bias by using statistical methods, which are described in the results section.

Coding and analyses

Each study included in the meta-analysis was read by two raters who coded the relevant effect sizes and study characteristics independently of each other. Disagreements were resolved by discussion. When the sample mean age was not reported we tried to estimate it based on other information in the article, for example, students’ grade levels. When information vital for coding was missing or was reported in an ambiguous way in an article, the corresponding author was asked to clarify by email.

Meta-analyses can only statistically combine effect sizes when their signs have the same meaning. We expected a positive relation between magnitude comparison and mathematical competence. All effects sizes were recoded prior to our analyses so that effects in line with our expectation had a positive sign and effect sizes that did not had a negative sign.

Age group was coded as below 6 years of age (i.e. before the onset of formal school instruction on whole numbers in most countries), between 6 and 9 years (i.e. during whole-number instruction in elementary school) or above 9 years (i.e. after whole-number instruction). The sample mean age was additionally coded as a continuous score, but only when the age range in a sample was five years or less.

Correlations are biased by measurement error. The lower the reliability of the measures is, the higher is the degree of random noise in the empirical data and the lower is the correlation between two variables, independent of the actual strength of the association of the underlying constructs. For this reason, when the reliability of the measures was reported in the original study, we modified the correlations using Spearman's correction for attenuation (Hunter & Schmidt, 2004, p. 96). When reliabilities of standardized tests were not reported, we took them from the test manuals. In cases of missing reliabilities the reported correlations were not corrected.

Effect sizes were weighted with the inverse of the standard error and combined using a random-effects model, which is adequate for inhomogeneous sets of effect sizes (e.g. Hedges & Vevea, 1998), in the statistical analysis software MPlus 7.1 (Muthén & Muthén, 1998–2012). The standard random-effects model requires the independence of all effect sizes. This was not the case in our data set because, for example, some studies reported several effect sizes found with the same sample. We solved this problem by implementing the random-effects model as a two-level model, in which effect sizes (level 1) were nested under independent samples (level 2). The use of two-level models for conducting meta-analyses is well understood from a statistical point of view and has been described as an elegant way of synthesizing clustered effect sizes (Hox, 2002, pp. 139–156; Van den Noortgate & Onghena, 2003). Correlations are the dependent variable of our meta-analysis. Sometimes, correlations are Fisher Z transformed before being used as interval-scale variables. We did not do this in our analyses, because newer studies found that meta-analyses on correlations lead to more accurate results without this transformation (Hunter & Schmidt, 2004, pp. 82–83). All reported confidence intervals (CI) are at the 95% level.

We entered all moderator variables as level-1 predictors of effect sizes into our two-level model, because their values can differ between effect sizes within independent samples. The only exception was age, which we modeled as level-1 predictor as well as level-2 predictor because our database included longitudinal studies, where age varies within independent samples, as well as cross-sectional studies, where age varies between independent samples.

Results

Study characteristics

The PsycINFO search returned 237 articles, 41 of which fulfilled the inclusion criteria (inter-rater agreement: 89% of the articles). Four additional articles (Brankaer, Ghesquiere & Smedt, 2014; Halberda *et al*., 2008; Lyons, Price, Vaessen, Blomert & Ansari, 2014; Reigosa-Crespo, González-Alemañy, León, Torres, Mosquera *et al*., 2013) were included in the meta-analysis because they were relevant, yet they were not listed in PsycINFO or were published shortly after we conducted the standardized literature search. The 45 articles (marked by an asterisk in the reference list) reported results from 79 independent samples with 284 relevant effect sizes and 17.201 participants in total. Inter-rater agreement was 99% for the coding of the effect sizes and 98% for the coding of study characteristics.

Of the 45 articles, 82% had been published in 2011 or later, indicating a rapid increase in research on the topic of this meta-analysis over recent years. Of the 284 effect sizes, 69% had been found with the non-symbolic magnitude comparison task and 31% with the symbolic magnitude comparison task. Of the studies using the non-symbolic task version, 32% also included numerosities in the subitizing range (i.e. one, two, and three dots), whereas 68% exclusively used numerosities greater three. Magnitude comparison measures were solution time (28%), solution rate (25%), Weber fraction (24%), distance effect with solution times (14%), ratio effect with solution times (4%), distance effect with solution rates (3%), and ratio effect with solution rates (2%). The competence measures were written arithmetic tasks (29%), curriculum-based tests (22%), mental arithmetic tasks (18%), tests of early mathematical ability (13%) and other tasks (18%) involving, for example, number decomposition, mathematical reasoning, or geometry. Reliability indices were available for 25% of magnitude comparison measures and 61% of mathematical competence measures. After correction, there was no statistically significant difference between effect sizes that were corrected for reliability and effect sizes that could not be corrected (p = .843, R^2 = .001, for reliabilities of magnitude comparison measures; p = .424, R^2 = .013 for reliabilities of competence measures). The majority of the participants (56%) were between 6 and 9 years old; 10% were younger and 34% older. The sample mean ages could be coded for 255 effect sizes and ranged from 3 to 35 years, even though some studies also included much older participants (Halberda *et al.*, 2012).

Overall effect for numerical magnitude comparison

The meta-analytic results are displayed in Table 1. The overall correlation between numerical magnitude comparison and mathematical competence according to the two-level random-effects model was $r = .278$ with a confidence interval ranging from .241 to .315. The H value of .418 indicates that 41.8% of the variance of the effect sizes is between the level-2 units, that is, between the 79 independent samples in our database. The H value is significantly different from zero with $p < .001$, which indicates a statistically significant amount of heterogeneity and implies that the magnitude comparison–competence relation is moderated by third variables.

Duval and Tweedie's trim-and-fill method indicates the absence of publication bias in our database. In this method, fictitious effect sizes are added to the left side of the effect size distribution until this distribution is symmetric and a new overall effect size can be computed for the symmetric distribution. Since our effect size distribution was already symmetric (see Figure 1), the trim-and-fill method left our results unchanged. In line with this, Rosenthal's fail-safe *N* had the value of 9101. Only if this extremely high number of unpublished studies with null results existed would the comparison–competence relation cease to be significant at the 5% level. Thus, the file-drawer problem is negligible in our case. The analyses also demonstrated that the results are not biased by an overly strong influence of specific samples. In a sensitivity analysis with the leave-one-out method the omission of a sample never changed the overall correlation by more than $\Delta r = \pm.008$ points with the exception of the study by Träff (2013). Leaving this study out would change the overall correlation by $\Delta r = -.012$. Träff found three effect sizes between 0.470 and 0.670 with 134 students from Grades 4 to 6 who solved the symbolic magnitude comparison task.

Table 1 *Number of effect sizes (k), correlation (r), 95% confidence interval, heterogeneity index (H) and p-value for the test of heterogeneity*

Analysis	*k*	Correlation *r*	lower 95% CI	upper 95% CI	Heterogeneity H	*p*
Overall	284	.278	.241	.315	.418	< .001
Task format						
Non-symbolic	195	.241	.198	.284	.478	< .001
Symbolic	89	.302	.243	.361	.429	.002
Age group						
Younger than 6 years	29	.305	.205	.402	.497	< .001
6 to 9 years	158	.283	.222	.344	.455	< .001
Older than 9 years	97	.280	.229	.331	.423	.016
Magnitude comparison measure						
Solution rate	72	.316	.245	.387	.670	.002
Solution time	79	.269	.216	.322	.511	.001
Distance effect (solution rate)[1]	7	−.081	−.185	.023	–	–
Distance effect (solution time)	41	.135	.080	.190	.085	.898
Ratio effect (solution rate)[1]	5	.140	−.080	.360	–	–
Ratio effect (solution time)	11	.142	.030	.254	.625	.105
Weber fraction	69	.315	.248	.382	.409	.092
Mathematical competence measure						
Early abilities (TEMA)	37	.413	.333	.493	.256	.016
Mental arithmetic	52	.378	.321	.435	.155	.162
Written arithmetic	81	.281	.189	.373	.483	< .001
Curriculum-based	62	.205	.138	.272	.316	.104
Other	52	.210	.159	.261	.235	.092

[1]Estimated by a one-level regression model due to the small number of sampling units.

Non-symbolic vs. symbolic comparison

The correlation between non-symbolic magnitude comparison and mathematical competence was $r = .241$, CI [.198, .284]. The correlation between symbolic magnitude comparison and mathematical competence was higher with $r = .302$, CI [.243, .361]. Task format (non-symbolic vs. symbolic) modeled as a level-1 predictor of the correlation coefficients was significantly related ($p < .001$) to the correlation coefficients and explained 9% of their variance. This indicates a small but statistically significant difference between correlations found with non-symbolic or with symbolic magnitude comparison.

Accounting for the moderating effect of task format did not reduce the heterogeneity of the effect sizes. As shown in Table 1, there was still considerable heterogeneity of effect sizes when only effects found with non-symbolic tasks were considered (H = .478; $p < .001$) or when only effects found with symbolic tasks were considered (H = .429; $p = .002$). This suggests the effects of further moderating variables. Whether the non-symbolic task version included or excluded stimuli in the subitizing range (i.e. one, two or three dots) did not significantly affect the effect size ($R^2 = .038$, $p = .200$).

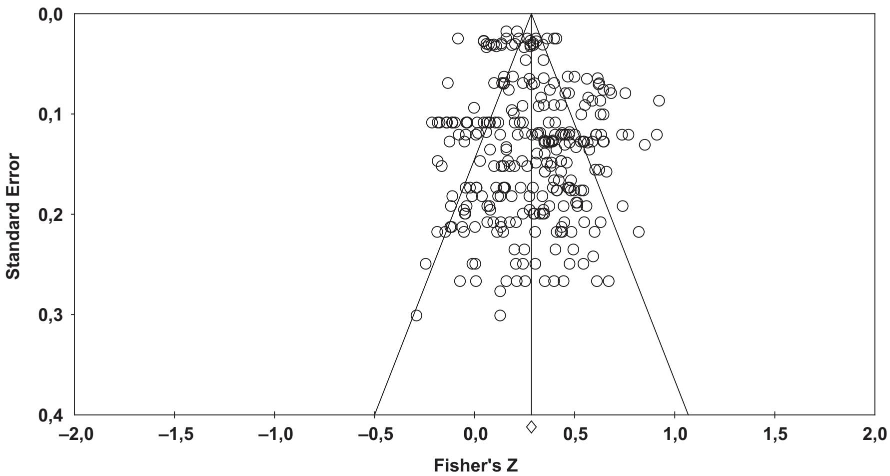

Figure 1 *Funnel plot of the 284 effect sizes. The inclined lines indicate the 95% confidence interval.*

Age and measures as potential moderators

The correlation between magnitude comparison and competence was similar for persons younger than 6 years (r = .305), persons aged 6 to 9 years (r = .283), and persons older than 9 years (r = .280). We modeled age group as dummy-coded level-1 predictor and dummy-coded level-2 predictor of effect sizes in separate analyses. Neither within-sample variations of age on level 1 (R^2 = 0.004; p = .796) nor between-sample variations of age on level 2 (R^2 = 0.007; p = .801) significantly moderated the effect sizes. In order to test for changes from adolescence to adulthood, we repeated our age group comparison with adults as a fourth group. The mean effect sizes were: 5 years or younger: .304 [CI .205, .405], 6–9 years: .283 [CI .222, .344], 10–17 years (adolescents): .311 [CI .235, .387] and 18 years or older (adults): .263 [CI .200, .326]. Again, age group coded as a set of dummy variables did not predict the effect sizes (all ps > .60, R^2 = .016).

In addition, we conducted meta-regressions with years of age coded as continuous variable, because continuous variables are more sensitive to gradual changes. Age as level-1 predictor of effect sizes had a small but significant effect (b = –.006, β = –.196, p = .013) and explained a variance proportion of R^2 = .038. In contrast, age as level-2 predictor was unrelated to the effect sizes (b = –.003, β = –.178, p = .105) and explained a variance proportion of R^2 = .032. Thus, the relationship between magnitude comparison and broader mathematical competence was very weakly moderated by age (see also Figure 2).

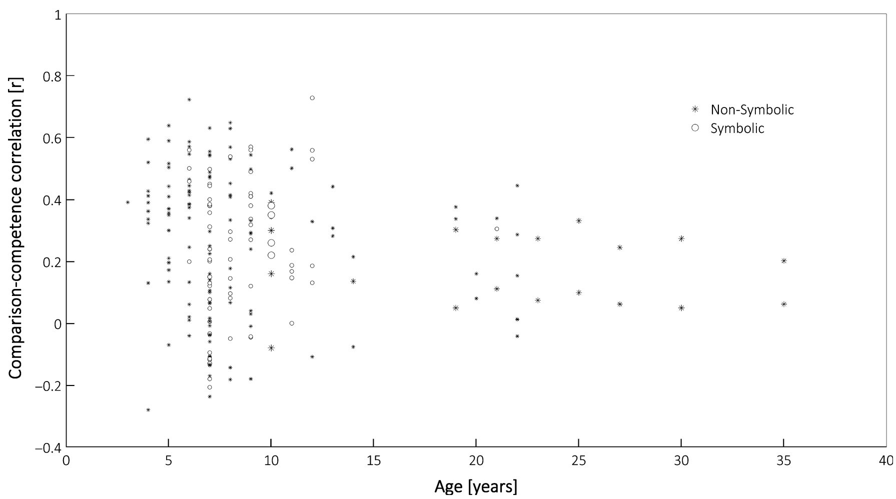

Figure 2 *Distribution of the effect sizes (Pearson correlations) by age and task format. Dot size is proportional to the sample size.*

As displayed in Table 1, the meta-analytic results differed strongly between magnitude comparison measures. The magnitude comparison measure as dummy-coded level-1 predictor explained a variance proportion of $R^2 = .142$, $p = .006$. Thus, the magnitude comparison–competence correlation was moderated by the type of magnitude comparison measure. The highest correlations with mathematical competence were found for solution rates and Weber fractions. The distance effect with solution rates and the ratio effect with solution rates were not correlated with mathematical competence, as is indicated by their confidence intervals, which include the zero. The H values in Table 1 demonstrate that effect sizes found only with solution rates or only with solution times still were heterogeneous and might be moderated by other variables.

The meta-analytic results also differed strongly between measures of mathematical competence. The mathematical competence measure as dummy-coded level-1 predictor explained a variance proportion of $R^2 = .138$, $p = .003$. The correlations were highest for early ability as assessed by the TEMA ($r = .413$) and for mental arithmetic ($r = .378$), both of which lay above the 95% confidence interval of the overall effect size in this meta-analysis. The correlations were lowest for curriculum-based measures ($r = .205$) and other measures ($r = .210$), both of which lay outside the 95% confidence interval found for the overall effect size in this meta-analysis (CI = .241–.315). The mean effect size found with mental arithmetic tasks ($r = .378$) was descriptively higher than the mean effect size found with written arithmetic tasks ($r = .281$), but this difference did not reach statistical significance ($R^2 = .130$; $p = .089$).

Exploratory analyses of moderator interactions

So far we have reported the statistical main effects of moderator variables on the comparison–competence association. This raises the question whether there are also interaction effects between the moderators. Analyzing this is not easy because effect sizes (here $k = 284$), and not individual persons (here $N = 17.201$), are the unit of analysis in meta-analyses. Thus, the statistical power of moderator-interaction tests in meta-analyses is usually limited. This was also the case in the present study. Obviously, the 284 effect sizes included did not allow us to delineate interaction effects of all 210 possible combinations of two task formats, three age groups, seven magnitude comparison measures, and five competence measures. We thus followed the advice of Hunter and Schmidt (2004, p. 426) and only conducted exploratory analyses of interaction effects. Specifically, we investigated the mean correlations separately for all combinations of task format and age group (see Table 2). We then repeated these analyses separately only for those comparison measures and those competence measures for which a statistically significant amount of heterogeneity had been found in our previous analyses (see last column of Table 1), which is suggestive of further moderating effects.

The correlations in Table 2 range from .161 to .489. Some cells were empty so that we could not compute these cell means. Most other cell counts were relatively low, which led to large confidence intervals. In some cases the cell counts were so low that the combined effect size for that cell had to be estimated based on a one-level regression model, which might underestimate within-study variance and, thus, might also underestimate the true range of the confidence interval (Van den Noortgate & Onghena, 2003).

Table 2 *Correlations, 95% CI in square brackets, and number of effect sizes in parentheses by age group, task format and measure*

Measure and task format	< 6 years	6–9 years	> 9 years
All measures			
Non-symbolic	.305 [.205, .405] ($k = 29$)	.224 [.150, .298] ($k = 92$)	.257 [.206, .308] ($k = 74$)
Symbolic	($k = 0$)	.279 [.210, .348] ($k = 66$)	.347 [.249, .445] ($k = 23$)
Solution rate as magnitude comparison measure			
Non-symbolic	.393 [.275, .511] ($k = 10$)	.244 [.117, .371] ($k = 34$)	.330 [.173, .487] ($k = 10$)
Symbolic	($k = 0$)	.250 [.115, .385] ($k = 14$)	.356[1] [.332, .380] ($k = 4$)
Solution time as magnitude comparison measure			
Non-symbolic	.357[1] [.257, .457] ($k = 4$)	.161 [.085, .237] ($k = 17$)	.165 [.110, .220] ($k = 22$)
Symbolic	($k = 0$)	.343 [.265, .421] ($k = 24$)	.446 [.344, .548] ($k = 12$)
Early abilities (TEMA) as mathematical competence measure			
Non-symbolic	.446[1] [.379, .513] ($k = 19$)	.454 [.323, .585] ($k = 18$)	($k = 0$)
Symbolic	($k = 0$)	($k = 0$)	($k = 0$)
Written arithmetic as mathematical competence measure			
Non-symbolic	($k = 1$)	.288 [.121, .455] ($k = 33$)	.251 [.096, .406] ($k = 20$)
Symbolic	($k = 0$)	.200 [.067, .333] ($k = 24$)	.489[1] [.377, .601] ($k = 3$)

[1]Estimated using a one-level model due to a small number of sampling units.

Figure 2 gives a more detailed account of age-related differences by displaying the distribution of effect sizes by age as a continuous dimension and by presentation of the task format. Each dot represents one of the 255 effect sizes for which age could be coded. The dot size is proportional to the sample size. In our view, the figure visualizes that there is no strong interaction between age, sample size and task format.

Overall, the results presented in this section demonstrated that there are no strong interaction effects between the moderating variables. Due to a lack of statistical power it remains an open question whether smaller interaction effects exist. Differences between single cells in Table 2 should not be over-interpreted.

Discussion

The current meta-analysis synthesized the published findings on the associations between non-symbolic or symbolic numerical magnitude comparison and mathematical competence. The strength of the association was r = .278, CI [.241, .315], averaged over all 284 effect sizes. In other words, the variance in magnitude comparison proficiency and the variance in mathematical competence overlap by about 8%. Whereas a variance proportion of this size is generally considered as indicating a weak effect, the effect size is remarkably high given how different magnitude comparison tasks and mathematical competence measures are both on a surface level and on the level of the concepts and strategies required for solving the respective task. As expected, there was a statistically significant amount of heterogeneity in the data, which indicates that the association between magnitude comparison and mathematical competence is moderated by third variables.

Non-symbolic vs. symbolic magnitude comparison

Task format moderated the association between magnitude comparison and mathematical competence. The average effect size found in our meta-analysis was significantly higher for symbolic magnitude comparison (.302, CI [.243, .361]) than for non-symbolic magnitude comparison (.241, CI [.198, .284]). This is in line with suggestions made by De Smedt *et al.* (2013) in their narrative review of the literature, who raised the possibility that the association between magnitude processing and broader mathematical competence might be more robust for studies with the symbolic magnitude processing tasks than for studies with the non-symbolic task. The current meta-analytic findings imply that symbolic magnitude processing is among the most eligible candidates to be targeted in intervention and diagnostic screening instruments for school-aged children at risk for mathematical difficulties. Before the onset of symbolic number knowledge in children, non-symbolic magnitude processing measures could be used to detect at-risk children, which could allow for earlier detection and possibly intervention.

The higher associations for symbolic than for non-symbolic comparison can be explained by the fact that measures of mathematical competence almost exclusively require the interpretation and transformation of texts and Arabic numerals, i.e. information presented in a symbolic from. Therefore, the symbolic comparison task is more similar to mathematical competence measures than the non-symbolic comparison task and might also involve more similar cognitive processes, such as symbol–referent mappings (Grabner, Ansari, Koschutnig, Reishofer & Ebner, 2013; Grabner, Reishofer, Koschutnig & Ebner, 2011).

It has recently been suggested that the association between non-symbolic numerical magnitude comparison and mathematical competence might be explained by more general non-numerical cognitive abilities, such as inhibitory control (Fuhs & McNeil, 2013; Gilmore, Attridge, Clayton, Cragg, Johnson *et al.*, 2013). A recent study contrasting the results of Fuhs and McNeil on the one hand and Gilmore and colleagues, on the other hand, found that after controlling for inhibitory control the non-symbolic comparison–competence correlation was weaker but still greater than zero (Keller & Libertus, 2015). In line with this, the meta-analysis by Chen and Li (2014) found that the overall effect size in studies controlling for general non-numerical cognitive abilities (.16 [.09, .24], k = 24) was significantly lower than the overall effect size in studies not controlling for them (.27 [.19, .35], k = 24). Both effect sizes were significantly different from zero, indicating that the correlation between magnitude comparison and mathematical competence cannot be entirely attributed to general non-numerical cognitive abilities. It is not yet known whether the same holds true for symbolic magnitude processing. So far, not many studies on non-symbolic or symbolic numerical magnitude comparison have included non-numerical cognitive variables in their analyses, and those who did varied considerably in the type of cognitive variable that was included, namely working memory (Träff, 2013), rapid automatized naming (Mazzocco, Feigenson & Halberda, 2011a), general processing speed (Bartelet *et al.*, 2014; Vanbinst, Ghesquière & Smedt, 2015), or attention (Libertus, Feigenson & Halberda, 2013a). In our meta-analysis, we did not control for these non-numerical cognitive variables. As the number of studies that include both numerical and non-numerical cognitive variables has been on the rise in current years, future meta-analyses should account for these non-numerical cognitive variables in greater detail.

Another important issue for further research concerns the degree of overlap between non-symbolic and symbolic magnitude processing in individuals. The dominant view assumes that symbolic representations are mapped onto non-symbolic ones, which are then further processed, in which case non-symbolic magnitude processing would be an important component of symbolic magnitude processing (see, e.g. Piazza, 2010, for a review). However, others (e.g. Bulthé, Smedt & Op de Beeck, 2014; Le Corre & Carey, 2007; Lyons, Ansari & Beilock, 2012) have argued that non-symbolic and symbolic processes develop independently from each other and might constitute different systems whose associations with mathematical competence might differ. The present meta-analysis, which focuses on relations of both tasks with competence, leaves this issue unresolved. There is a need for future studies to longitudinally follow up children's development of non-symbolic and symbolic magnitude processing and how their association changes over time.

For the non-symbolic magnitude comparison task, the result of the current meta-analysis is very close to the results of the two previous meta-analyses, which were .20, CI [.14, .26] (Chen & Li, 2014) and .22, CI [.20, .25] (Fazio *et al.*, 2014). The minor differences between the results are likely due to three small methodological differences between the three meta-analyses. First, the previous meta-analyses had coded only one effect size from each sample (Chen and Li 47 effect sizes; Fazio and colleagues 34 effect sizes) whereas our multilevel model allowed us to include all effect sizes from all samples, resulting in a total of 195 effect sizes for non-symbolic magnitude comparison. Second, we used an attenuation correction for the non-perfect reliabilities of the measures, which was not done in the previous two meta-analyses. Finally, Chen and Li included correlations controlled for third variables whenever these were available, in order to deconfound the comparison–competence relation from the influences of third variables. We decided not to do this in the present study, because the combination of correlations controlled for conceptually different variables is hard to interpret. Despite these methodological differences, the findings from the three meta-analyses differ only slightly and the confidence intervals overlap to a large extent. This demonstrates the robustness of the findings against methodological decisions made by the three groups of authors.

Age as moderator

The association between numerical magnitude comparison and mathematical competence was only weakly moderated by age, as indicated by a very small decrease of the magnitude–competence correlation with age. Neither our age group comparisons nor regressions with age as continuous level-2 predictor were able to detect this effect. Only a regression with age as continuous level-1 predictor of effect sizes indicated a weak negative relation. This pattern of results is plausible, because the latter analysis has a higher statistical power than the former three analyses for two reasons. Analyses with age as continuous predictor have higher sensitivity to gradual age-related changes than age group comparisons; and methodological differences between studies are confounded with age effects on level 2 (i.e. studies differ in method and participant age), but not on level 1 (i.e. the same method is used in all age groups of the same study).

The finding that the comparison–competence correlation decreases only slightly with participant age might be surprising, because the link between magnitude comparison and the content of mathematics instruction beyond the elementary-school years is far from obvious and is subject to ongoing discussions (Rips, Bloomfield & Asmuth, 2008; Szücs, Soltész & Goswami, 2009). However, our findings are consistent with the results of the previous two meta-analyses (Chen & Li, 2014; Fazio *et al.*, 2014) and of the only large-scale study ($N \geq 10{,}000$) on the lifespan development of non-symbolic magnitude comparison so far (Halberda *et al.*, 2012), all of which either found no moderating effect of age at all or only a relatively small moderating effect. The absence of an age effect might indicate that the association between magnitude processing and mathematical competence is stable over the lifespan. Alternatively, it might be that this association is mainly driven by an effect of numerical magnitude processing on early mathematics development, which then cascades into future mathematics development throughout life. Most of the studies included in this meta-analysis were cross-sectional, and there is a need for longitudinal data to investigate this issue. These future studies will have to test whether any causal relations between magnitude processing and math competence are stable over the lifespan or whether these relations diminish while their academic and motivational consequences remain over the lifespan.

Implications of the findings

Our meta-analytic findings have at least three implications, the first of which concerns the measures used in this field of research. The choice of measures strongly influences the correlation between comparison and competence, even more so than the choice of a non-symbolic or a symbolic task format. In our analyses, differences between measures of magnitude comparison skills explained about 14% of the variance of effect sizes found with these measures. Differences between measures of mathematical competence also explained about 14% of the variance of effect sizes. In contrast, task format explained only 9% of the variance. For the magnitude comparison measures, solution rates, solution times and Weber fractions led to substantially stronger effects than the use of the distance effect or of the ratio effect. One explanation for the lower values found with distance effects and ratio effects could be the fact that they are difference scores (or change rates). Difference scores and change rates tend to have lower reliabilities than sum scores at least under some circumstances. Their usefulness is the subject of an ongoing methodological debate (May & Hittner, 2003; Williams & Zimmerman, 1996). Our results are also in line with findings of Price *et al*. (2012) who found Weber fractions to be more reliable than ratio effects. We did correct for non-perfect reliabilities of magnitude comparison measures in our meta-analysis, but only for those 25% of studies that reported the reliabilities of the magnitude comparison measures. Thus, different reliabilities can at least partly explain the different effect sizes found with different measures (see Dietrich, Huber & Nuerk, 2015, for a review of magnitude comparison measures and their reliabilities). An additional explanation for the substantial correlation between accuracy measures of magnitude processing skills and mathematical competence tests is that both measures are based on solution rates and, thus, indicate the breadth of a person's numerical or mathematical knowledge. In contrast, ratio and distance effect quantify the precision of mental magnitude representations and, thus, assess a slightly different construct than competence tests (Dietrich *et al*., 2015).

For mathematical competence measures, the strongest effects were found with the TEMA and mental arithmetic as compared to written arithmetic, curriculum-based measures and various other tasks. Potential explanations for these differences are that the TEMA draws on numerical abilities that are conceptually close to magnitude comparison, such as counting or informal calculation (e.g. with fingers) (Libertus, Feigenson & Halberda, 2013b). It is also important to note that one of the eight areas of the TEMA-3 involves number comparison, although not all children receive all of the six comparison items. So at least a small part of the association might also be attributed to surface similarities between the test and the magnitude comparison task. In mental arithmetic, learners need to mentally represent the magnitudes of the two addends and the magnitude of the sum (DeStefano & LeFevre, 2004). A good understanding of the numbers in terms of their magnitudes can be helpful for choosing an efficient calculation strategy (e.g. Peters, Smedt, Torbeyns, Verschaffel & Ghesquière, 2014). In contrast, written arithmetic merely requires learners to combine the two addends digit by digit by following a standard algorithm, and there is no need for mentally representing the magnitudes of the addends and the sum (Linsen, Verschaffel, Reynvoet & Smedt, 2015). Curriculum-based measures typically include a broad variety of mathematical skills, not all of which depend on numerical magnitude processing (Schneider *et al*., 2009).

The second implication of our findings is the need for studies with experimental designs which allow tests of causal hypotheses about the relations between magnitude processing and mathematical competence. As our meta-analysis shows, there is an abundance of correlational studies. These studies leave unanswered whether numerical magnitude processing causally affects mathematical competence. There are some intervention studies in which numerical magnitude comparison skills have been successfully trained (see De Smedt *et al*., 2013, for a review). Some of these studies have indicated small but positive transfer effects to untrained mathematical tasks (e.g. Obersteiner, Reiss & Ufer, 2013; Ramani & Siegler, 2011; Siegler & Ramani, 2009). Park and Brannon (2013, 2014) showed that training adults' approximate non-symbolic arithmetic skills improved their performance on a symbolic two- and three-digit addition and subtraction test. Hyde, Khanum and Spelke (2014) reported that brief training on non-symbolic numerical magnitude comparison improved 6–7-year-olds' symbolic addition skills. These results suggest a causal influence of numerical magnitude processing on broader mathematical competence. However, other studies have failed to observe such transfer effects (Räsänen, Salminen, Wilson, Aunio & Dehaene, 2009; Wilson, Dehaene, Dubois & Fayol, 2009), and some of the successful studies have been conducted with children from low-income backgrounds or children with low numeracy. So the generalizability of findings on causal relations is not yet fully clear. It is also unclear whether, in addition to the possible influence of comparison on competence, there is also a causal influence in the opposite direction, so that gains in mathematical competence cause improvements in magnitude comparison skills. This hypothesis is supported by studies with adults which found positive effects of mathematics instruction on non-symbolic number comparison abilities (Nys, Ventura, Fernandes, Querido, Leybaert *et al*., 2013; Piazza, Pica, Izard, Spelke & Dehaene, 2013). Future intervention studies should systematically test these alternative causal explanations for the correlations found in our meta-analysis.

Finally, Chen and Li (2014) have argued that, given the relatively small correlation between non-symbolic magnitude comparison and mathematical competence, many studies in this field of research were severely underpowered because they lacked the large sample sizes needed to detect small effects. Our analyses demonstrate that this issue is somewhat less problematic for the symbolic magnitude comparison task. According to power analyses conducted with G*Power 3 (Faul, Erdfelder, Lang & Buchner, 2007), given the effect sizes found in our meta-analyses, a critical alpha error level of 5%, a statistical power of 80%, and one-sided testing, 102 participants are needed to detect the correlation between non-symbolic magnitude comparison and math competence, but only 64 participants are needed to detect the correlation between symbolic magnitude comparison and mathematical competence.

Acknowledgements

Bert De Smedt is partially supported by grant G.0946.12 of the Research Foundation Flanders (FWO), Belgium, and grant GOA 2012/010 of the Research Fund KU Leuven, Belgium.

References

* Agrillo, C., Piffer, L., & Adriano, A. (2013). Individual differences in non-symbolic numerical abilities predict mathematical achievements but contradict ATOM. *Behavioral and Brain Functions*, **9** (26). doi: 10.1186/1744-9081-9-26

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature Reviews Neuroscience*, **9**, 278–291.

* Bartelet, D., Vaessen, A., Blomert, L., & Ansari, D. (2014). What basic number processing measures in kindergarten explain unique variability in first-grade arithmetica proficiency? *Journal of Experimental Child Psychology*, **117**, 12–28. doi:10.1016/j.jecp.2013.08.010

Berch, D.B. (2005). Making sense of number sense: implications for children with mathematical disabilities. *Journal of Learning Disabilities*, **38** (4), 333–339.

* Bonny, J.W., & Loureneo, S.F. (2013). The approximate number system and its relation to early math achievement: evidence from the preschool years. *Journal of Experimental Child Psychology*, **114**, 375–388. doi:10.1 016/j.jecp.2012.09.015

Booth, J.L., & Newton, K.J. (2012). Fractions: could they really be the gatekeeper's doorman? *Contemporary Educational Psychology*, **37**(4), 247–253. doi: 10.1016/j.cedpsych.2012.07.001

* Brankaer, C., Ghesquiere, P., & Smedt, B.D. (2014). Children's mapping between non-symbolic and symbolic numerical magnitudes and its association with timed and untimed tests of mathematics achievement. *PLOS ONE*, **9** (4), e93565. doi: 10.1371/journal.pone.0093565

* Bugden, S., & Ansari, D. (2011). Individual differences in children's mathematical competence are related to the intentional but not automatic processing of arabic numerals. *Cognition*, **118**, 32–44. doi: 10.1016/j.cognition.2010.09.005

* Bugden, S., Price, G.R., McLean, D.A., & Ansari, D. (2012). The role of the left intraparietal sulcus in the relationship between symbolic number processing and children's arithmetic competence. *Developmental Cognitive Neuroscience*, **2**, 448–457. doi:10.1016/j.dcn.2012.04.001

Bulthé, J., Smedt, B.D., & Op de Beeck, H.P. (2014). Format-dependent representations of symbolic and non-symbolic numbers in the human cortex as revealed by multi-voxel pattern analyses. *NeuroImage*, **87**, 311–322. doi:10.1016/j.neuroimage.2013.10.049

* Butterworth, B., Zorzi, M., Girelli, L., & Jonckheere, A.R. (2001). Storage and retrieval of addition facts: the role of number comparison. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, **54** (4), 1005–1029. doi:10.1080/713756007

Cantlon, J.F. (2012). Math, monkeys, and the developing brain. *Proceedings of the National Academy of Sciences, USA*, **109**, 10725–10732.

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

* Castronovo, J., & Göbel, S.M. (2012). Impact of high mathematics education on the number sense. *PLoS ONE*, **7** (4), e33832. doi:10.1371/journal.pone.0033832

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. doi:10.1016/j.actpsy.2014.01.016

Cohen Kadosh, R., Henik, A., Rubinsten, O., Mohr, H., Dori, H. *et al*. (2005). Are numbers special? The comparison systems of the human brain investigated by fMRI. *Neuropsychologia*, **43**, 1238–1248. doi:10.1016/j.neuropsychologia.2004.12.017

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, **16** (3), 626–641.

De Smedt, B., Noël, M.-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, **2** (2), 48–55. doi:10.1016/j.tine.2013.06.001

* De Smedt, B., Verschaffel, L., & Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *Journal of Experimental Child Psychology*, **103**, 469–479.

DeStefano, D., & LeFevre, J.A. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology*, **16** (3), 353–386. doi:10.1080/09541440244000328

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

* Dos Santos, F.H., Silva, P.A.D., Ribeiro, F.S., Dias, A.L.R.P., Frigério, M.C. *et al*. (2012). Number processing and calculation in Brazilian children aged 7–12 years. *Spanish Journal of Psychology*, **15** (2), 513–525. doi:10.5209/rev_SJOP.2012.v15.n2.38862

Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K., Huston, A.C. *et al*. (2007). School readiness and later achievement. *Developmental Psychology*, **43** (6), 1428–1446. doi:10.1037/0012-1649.43.6.1428

* Durand, M., Hulme, C., Larkin, R., & Snowling, M. (2005). The cognitive foundations of reading and arithmetic skills in 7- to 10-year-olds. *Journal of Experimental Child Psychology*, **91**, 113–136.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, **39** (2), 175–191.

* 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. doi:10.1016/j.jecp.2014.01.013

Feigenson, L., Libertus, M.E., & Halberda, J. (2013). Links between the intuitive sense of number and formal mathematics ability. *Child Development Perspectives*, **7** (2), 74–79. doi:10.1111/cdep.12019

Ferguson, C.J., & Brannick, M.T. (2011). Publication bias in psychological science: prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods*, **17** (1), 120–128. doi:10.1037/a0024445

* 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. doi:10.1111/desc.12013

* Gebuis, T., & van der Smagt, M.J. (2011). False approximations of the approximate number system? *PLoS ONE*, **6** (10), doi:10.1371/journal.pone.0025405

Gerardi, K., Goette, L., & Meier, S. (2013). Numerical ability predicts mortgage default. *Proceedings of the National Academy of Sciences, USA*, **110** (28), 11267–11271. doi:10.1073/pnas.1220568110

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S. *et al*. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLOS ONE*, **8** (6), e67374. doi:10.1371/journal.pone.0067374

Gilmore, C., Attridge, N., & Inglis, M. (2011). Measuring the approximate number system. *Quarterly Journal of Experimental Psychology*, **64** (11), 2099–2109. doi:10.1080/17470218.2011.574710

Gilmore, C., McCarthy, S.E., & Spelke, E. (2007). Symbolic arithmetic without instruction. *Nature*, **447**, 589–591.

Ginsburg, H.P., & Baroody, A.J. (2003). *Test of early numeracy mathematics ability* (3rd edn.). Austin, TX: Pro-Ed.

Grabner, R.H., Ansari, D., Koschutnig, K., Reishofer, G., & Ebner, F. (2013). The function of the left angular gyrus in mental arithmetic: evidence from the associative confusion effect. *Human Brain Mapping*, **34** (5), 1013–1024. doi:10.1002/hbm.21489

Grabner, R.H., Reishofer, G., Koschutnig, K., & Ebner, F. (2011). Brain correlates of mathematical competence in processing mathematical representations. *Frontiers in Human Neuroscience*, **5**, 130. doi:10.3389/fnhum.2011.00130

* Guillaume, M., Nys, J., Mussolin, C., & Content, A. (2013). Differences in the acuity of the approximate number system in adults: the effect of mathematical ability. *Acta Psychologica*, **144**, 506–512. doi:10.1016/j.actpsy.2013.09.001

* Halberda, J., Lya, R., Wilmer, J.B., Naiman, D.Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive internet-based sample. *Proceedings of the National Academy of Sciences, USA*, **109** (28), 11116–11120. doi:10.1073/pnas.1200196109

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

Hedges, L.V., & Vevea, J.L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, **3** (4), 486–504.

Holloway, I.D., & Ansari, D. (2008). Domain-specific and domain-general changes in children's development of number comparison. *Developmental Science*, **11** (5), 644–649. doi:10.1111/j.1467-7687.2008.00712.x

Holloway, I.D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: the numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology*, **103**, 17–29. doi:10.1016/j.jecp.2008.04.001

Hox, J. (2002). *Multilevel analysis: Techniques and applications.* Mahwah, NJ: Erlbaum.

Hunter, J.E., & Schmidt, F.L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd edn.). Thousand Oaks, CA: Sage.

Hyde, D.C., Khanum, S., & Spelke, E.S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition*, **131**, 92–107. doi:10.1016/j.cognition.2013.12.007

* Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathemtics achievement: but only in children. *Psychonomic*

Bulletin & Review, **18** (6), 1222–1229. doi:10.3758/s13423-011-0154-1

Inglis, M., & Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychologica*, **145**, 147–155. doi: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**, 685. doi:10.3389/fpsyg.2015.00685

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: an investigation of the conceptual sources of the verbal counting principles. *Cognition*, **105** (2), 395–438. doi:10.1016/j.cognition.2006.10.005

Libertus, M.E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science*, **14** (6), 1292–1300. doi: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.

* 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**, 829–838.

* Libertus, M.E., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica*, **141** (3), 373–379. doi:10.1016/j.actpsy.2012.09.009

Linsen, S., Verschaffel, L., Reynvoet, B., & De Smedt, B. (2015). The association between numerical magnitude processing and mental versus algorithmic multi-digit subtraction in children. *Learning and Instruction*, **35**, 42–50. doi:10.1016/j.learninstruc.2014.09.003

* Lonnemann, J., Linkersdörfer, J., Hasselhorn, M., & Lindberg, S. (2011). Symbolic and non-symbolic distance effects in children and their connection with arithmetic skills. *Journal of Neurolinguistics*, **24**, 583–591. doi: 10.1016/jJneuroling.2011.02.004

* Lourenco, S.F., Bonny, J.W., Fernandez, E.P., & Rao, S. (2012). Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence. *PNAS Early Edition*. doi: 10.1073/pnas.1207212109

Lyons, I.M., Ansari, D., & Beilock, S.L. (2012). Symbolic estrangement: evidence against a strong association between numerical symbols and the quantities they represent. *Journal of Experimental Psychology: General*, **141**, 635–641.

* Lyons, I.M., & 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

* Lyons, I.M., Price, G.R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1–6. *Developmental Science*, **17** (5), 714–726. doi:10.1111/desc.12152

May, K., & Hittner, J.B. (2003). On the relation between power and reliability of difference scores. *Perceptual and Motor Skills*, **97**, 905–908. doi:10.2466/pms.2003.97.3.905

* Mazzocco, M.M.M., Feigenson, L., & Halberda, J. (2011a). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development*, **82** (4), 1224–1237. doi:10.1111/j.1467-8624.2011.01608.x

* Mazzocco, M.M.M., Feigenson, L., & Halberda, J. (2011b). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE*, **6** (9), e23749. doi:10.1371/journal.pone.0023749

Moyer, R.S., & Landauer, T.K. (1967). Time required for judgments of numerical inequality. *Nature*, **215**, 1519–1520.

Muthén, B.O., & Muthén, L.K. (1998–2012). *Mplus user's guide* (7th edn.). Los Angeles, CA: Muthen & Muthen.

* Nys, J., & Content, A. (2012). Judgement of discrete and continuous quantity in adults: number counts! *Quarterly Journal of Experimental Psychology*, **65** (4), 675–690. doi:10.1080/17470218.2011.619661

Nys, J., Ventura, P., Fernandes, T., Querido, L., Leybaert, J. *et al*. (2013). Does math education modify the approximate number system? A comparison of schooled and unschooled adults. *Trends in Neuroscience and Education*, **2** (1), 13–22. doi:10.1016/j.tine.2013.01.001

Obersteiner, A., Reiss, K., & Ufer, S. (2013). How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. *Learning and Instruction*, **23**, 125–135. doi:10.1016/j.learninstruc.2012.08.004

Park, J., & Brannon, E.M. (2013). Training the approximate number system improves math proficiency. *Psychological Science*, **24** (10), 2013–2019. doi:10.1177/0956797613482944

Park, J., & Brannon, E.M. (2014). Improving arithmetic performance with number sense training: an investigation of underlying mechanism. *Cognition*, **133**, 188–200. doi:10.1016/j.cognition.2014.06.011

Peters, G., Smedt, B.D., Torbeyns, J., Verschaffel, L., & Ghesquière, P. (2014). Subtraction by addition in children with mathematical learning disabilities. *Learning and Instruction*, **30**, 1–8. doi:10.1016/j.learn.instruc.2013.11.001

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

Piazza, M., Pica, P., Izard, V., Spelke, E.S., & Dehaene, S. (2013). Education enhances the acuity of the nonverbal approximate number system. *Psychological Science*, **24** (6), 1037–1043. doi:10.1177/0956797612464057

* Praet, M., Titeca, D., Ceulemans, A., & Desoete, A. (2013). Language in the prediction of arithmetics in kindergarten and grade 1. *Learning and Individual Differences*, **27**, 90–96. doi: 10.1016/j.lindif.2013.07.003

* Price, G.R., Palmer, D., Battista, C., & Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults.

Acta Psychologica, **140**, 50–57. doi: 10.1016/j.actpsy.2012.02.008

Ramani, G.B., & Siegler, R.S. (2011). Reducing the gap in numerical knowledge between low- and middle-income preschoolers. *Journal of Applied Developmental Psychology*, **32**, 146–159. doi:10.1016/j.appdev.2011.02.005

Räsänen, P., Salminen, J., Wilson, A.J., Aunio, P., & Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. *Cognitive Development*, **24** (4), 450–472. doi:10.1016/j.cogdev.2009.09.003

* Reigosa-Crespo, V., González-Alemañy, E., León, T., Torres, R., & Mosquera, R. *et al*. (2013). Numerical capacities as domain-specific predictors beyond early mathematics learning: a longitudinal study. *PLoS ONE*, **8** (11), e79711. doi: 10.1371/journal.pone.0079711

Reyna, V.F., Nelson, W.L., Han, P.K., & Dieckmann, N.F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, **135** (6), 943–973. doi:10.1037/a0017327

Rips, L.J., Bloomfield, A., & Asmuth, J. (2008). From numerical concepts to concepts of number. *Behavioral and Brain Sciences*, **31** (6), 623–642. doi:10.1017/S0140525X08005566

Ritchie, S.J., & Bates, T.C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological Science*, **24** (3), 1301–1308. doi:10.1177/0956797612466268

* Rousselle, L., & Noël, M.-P. (2008). The development of automatic numerosity processing in preschoolers: evidence for numerosity-preceptual interference. *Developmental Psychology*, **44** (2), 544–560. doi:10.1037/0012-1649.44.2.544

* Sasanguie, D., Van den Bussche, E., & Reynvoet, B. (2012a). Predictors for mathematics achievement? Evidence from a longitudinal study. *Mind, Brain and Education*, **6** (3), 119–128. doi: 10.1111/j.1751-228X.2012.01147.x

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

* Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012b). Association between basic numerical abilities and mathematics achievement. *British Journal of Developmental Psychology*, **30**, 344–357. doi: 10.1111/j.2044-835X.2011.02048.x

Schley, D.R., & Peters, E. (2014). Assessing 'economic value': symbolic-number mappings predict risky and riskless valuations. *Psychological Science*, **25** (3), 753–761. doi:10.1177/0956797613515485

* Schneider, M., Grabner, R.H., & Paetsch, J. (2009). Mental number line, number line estimation, and mathematical achievement: their interrelations in grades 5 and 6. *Journal of Educational Psychology*, **101** (2), 359–372.

Schneider, M., Rittle-Johnson, B., & Star, J.R. (2011). Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge. *Developmental Psychology*, **46**, 1525–1538.

Siegler, R.S., & Ramani, G.B. (2009). Playing linear number board games – but not circular ones – improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology*, **101** (3), 545–560. doi:10.1037/a0014239

* Soltész, F., Szücs, D., & Szücs, L. (2010). Relationships between magnitude representation, counting and memory in 4- to 7-year-old children: a developmental study. *Behavioral and Brain Functions*, **6**, 13. doi: 10.1186/1744-9081-6-13

* Starke, M., Kiechl-Kohlendorfer, U., Kucian, K., Peglow, U.P., & Kremser, C. *et al*. (2013). Brain structure, number magnitude processing, and math proficiency in 6- to 7-year-old children born prematurely: a voxel-based morphometry study. *NeuroReport*, **24**, 419–424. doi: 10.10971WNR.Ob013e32836140ed

* Starr, A., Libertus, M.E., & Brannon, E.M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences, USA*, **110** (45), 18116–18120. doi: 10.1073/pnas.1302751110

Szücs, D., Soltész, F., & Goswami, U. (2009). Beyond format-specificity: is analogue magnitude really the core abstract feature of the cultural number representation? *Behavioral and Brain Sciences*, **32** (3/4), 352–353. doi:doi:10.1017/S0140525X09990884

* Träff, U. (2013). The contribution of general cognitive abilities and number abilities to different aspects of mathematics in children. *Journal of Experimental Child Psychology*, **116** (2), 139–156. doi: 10.1016/j.jecp.2013.04.007

Van den Noortgate, W., & Onghena, P. (2003). Multilevel meta-analysis: a comparison with traditional meta-analytical procedures. *Educational and Psychological Measurement*, **63**, 765–790. doi:10.1177/0013164402251027

Van Opstal, F., Gevers, W., De Moor, W., & Verguts, T. (2008). Dissecting the symbolic distance effect: comparison and priming effects in numerical and nonnumerical orders. *Psychonomic Bulletin & Review*, **15** (2), 419–425. doi:10.3758/PBR.15.2.419

* Vanbinst, K., Ghesquière, P., & De Smedt, B. (2012). Numerical magnitude representations and individual differences in children's arithmetic strategy use. *Mind, Brain and Education*, **6** (3), 129–136. doi:10.1111/j.1751-228X.2012.01148.x

Vanbinst, K., Ghesquière, P., & De Smedt, B. (2015). Does numerical processing uniquely predict first graders' future development of single-digit arithmetic? *Learning and Individual Differences*, **37**, 153–160. doi:10.1016/j.lindif.2014.12.004

* VanDerHeyden, A.M., Broussard, C., Snyder, P., George, J., & Lafleur, S.M. *et al*. (2011). Measurement of kindergartners' understanding of early mathematical concepts. *School Psychology Review*, **40** (2), 296–306.

* Wei, W., Lu, H., Zhao, H., Chen, C., Dong, Q. *et al*. (2012). Gender differences in children's arithmetic performance are accounted for by gender differences in language abilities. *Psychological Science*, **23** (3), 320–330. doi:10.1177/0956797611427168

Williams, R.H., & Zimmerman, D.W. (1996). Are simple gain scores obsolete? *Applied Psychological Measurement*, **20** (1), 59–69. doi:10.1177/014662169602000106

Wilson, A.J., Dehaene, S., Dubois, O., & Fayol, M. (2009). Effects of an adaptive game intervention on accessing number sense in low-socioeconomic-status kindergarten children. *Mind, Brain, and Education*, **3** (4), 224–234. doi:10.1111/j.1751-228X.2009.01075.x

Received: 1 December 2014
Accepted: 29 August 2015