Cognition, emotion, and arithmetic in primary school: A cross-cultural investigation

Maja Rodic¹,², Jiaxin Cui³, Sergey Malykh¹,⁴, Xinlin Zhou³, Elena I. Gynku¹, Elena L. Bogdanova⁵, Dina Y. Zueva¹, Olga Y. Bogdanova⁵ and Yulia Kovas¹,²*

¹Laboratory for Cognitive Investigations and Behavioural Genetics, Tomsk State University, Tomsk, Russia
²InLab, Department of Psychology, Goldsmiths, University of London, UK
³State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China
⁴Psychological Institute, Russian Academy of Education, Moscow, Russia
⁵Unit of General and Educational Psychology, Psychology Department, Tomsk State University, Tomsk, Russia

The study investigated cross-cultural differences in variability and average performance in arithmetic, mathematical reasoning, symbolic and non-symbolic magnitude processing, intelligence, spatial ability, and mathematical anxiety in 890 6- to 9-year-old children from the United Kingdom, Russia, and China. Cross-cultural differences explained 28% of the variance in arithmetic and 17.3% of the variance in mathematical reasoning, with Chinese children outperforming the other two groups. No cross-cultural differences were observed for spatial ability and mathematical anxiety. In all samples, symbolic magnitude processing and mathematical reasoning were independently related to early arithmetic. Other factors, such as non-symbolic magnitude processing, mental rotation, intelligence, and mathematical anxiety, produced differential patterns across the populations. The results are discussed in relation to potential influences of parental practice, school readiness, and linguistic factors on individual differences in early mathematics.

Statement of contribution
What is already known on this subject?

• Cross-cultural differences in mathematical ability are present in preschool children.
• Similar mechanisms of mathematical development operate in preschool children from the United Kingdom, Russia, and China.
• Tasks that require understanding of numbers are best predictors of arithmetic in preschool children.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

*Correspondence should be addressed to Yulia Kovas, Department of Psychology, Goldsmiths, University of London, London SE1 4 6NW, UK (email: y.kovas@gold.ac.uk).
Maja Rodic and Jiaxin Cui contributed equally to this work.

DOI:10.1080/02672496.2018.1425003
What does this study add?

- Cross-cultural differences in mathematical ability become greater with age/years of formal education.
- Similar mechanisms of mathematical development operate in early primary school children from the United Kingdom, Russia, and China.
- Symbolic number magnitude and mathematical reasoning are the main predictors of arithmetic in all three populations.

Early mathematics achievement has been found to be important for later mathematical development (Entwisle & Alexander, 1989). Children who fall behind in mathematics in early years of education tend to fall even further behind over time, with lower performing children not being able to keep with the pace in acquiring mathematical information (Aunola, Leskinen, Lerkkanen, & Nurmi, 2004; Desoete & Grégoire, 2006). This effect is present even after controlling for the influence of the family background, child’s intelligence, and reading ability (Starkey & Klein, 2007). The opposite pattern – the Matthew effect – suggests accumulated advantage, where children who perform well early on show even bigger advantage over time. Studying mathematical ability cross-culturally is useful for understanding the sources of individual differences in mathematical competence within any population, as well as for identification of universal features of mathematical development.

The average mathematical advantage of children of all ages from Asian–Pacific countries has been very well established (Gonzales et al., 2008; Imbo & LeFevre, 2009; Mullis, Martin, & Foy, 2008; OECD, 2010). This advantage seems to exist even before the beginning of formal schooling (Huntsinger, Jose, Liaw, & Ching, 1997; Rodic et al., 2015; Stevenson & Stigler, 1992) and is present throughout the school years (Geary, Bow-Thomas, Fan, & Siegler, 1993; Leung, 2006; Miura, 1987).

Explanations for the observed cross-cultural differences in mathematics performance include linguistic factors, such as transparency of a number system and speed of pronunciation of numbers (Dehaene, 1997); parental support (Chao, 2001; Huntsinger et al., 1997); educational systems (Stevenson, Chen, & Lee, 1993); cultural beliefs (Campbell & Xue, 2001); school readiness (Miller, Kelly, & Zhou, 2005); and genetic differences among populations. Our previous research suggested that socio-demographic and linguistic factors contribute to cross-cultural differences in performance of preschoolers (Rodic et al., 2015).

Cross-cultural differences in non-cognitive domains have also been found (e.g., Lee, 2009). Typically, students from Asian countries show a lower level of self-efficacy and self-concept and higher levels of mathematical anxiety than participants from Western countries (Lee, 2009).

In addition to exploring cross-cultural differences in variation and average levels of performance, it is important to understand whether differences exist in interrelationships among the different domains across the populations. In particular, it is necessary to explore whether different aspects of cognition, motivation, and emotion regulation (such as maths anxiety) are equally strongly related to mathematical skills in different populations and at different ages. Previous research suggested high degree of similarity across cultures in these interrelationships showing that, for example, symbolic magnitude processing task is related to early arithmetic in the UK, Russian, Chinese, Kyrgyz, and Dungan 5- to 7-year-old children (Rodic et al., 2015).
Overall, a wealth of research in different populations shows that ability to process numerical magnitudes of symbolic numbers is positively related to development of arithmetic skills and mathematics achievement (De Smedt, Verschaffel, & Ghesquiere, 2009; Gilmore, McCarthy, & Spelke, 2010; Holloway & Ansari, 2009; Kolkman, Kroesbergen, & Leseman, 2013; Zhang, Chen, Liu, Cui, & Zhou, 2016). This relationship may be stronger in older children, as practice with mathematical operations and exposure to numbers increases with age (Clements & Sarama, 2007; Dehaene, 1997). However, culture-specific influences may increase or decrease the magnitude of the interrelationships over time. For example, previous research has suggested cross-cultural differences in associations between maths performance and numerical factors in preschool children (Rodic et al., 2015). It is necessary to replicate these differences and to investigate whether they remain significant in children at the beginning of formal maths education.

Not all previous research has been consistent in finding substantial positive associations between mathematical skills and basic magnitude processing. For example, number comparison performance in 9-year-old children showed only a modest association with mathematical skills (Szucs, Devine, Soltesz, Nobes, & Gabriel, 2014). Another study has not found significant relationships between performance on single-digit number comparison and number reasoning, arithmetic learning, and advanced mathematics in adults (Wei, Yuan, Chen, & Zhou, 2012). Therefore, further research is needed to identify the sources of these inconsistencies.

Finally, intelligence and other cognitive factors typically explain no more than half of the variability in mathematical variation. The other half of the variance is related to other factors, including motivation, emotion, and health (Krapohl et al., 2014). Of these factors, mathematical anxiety has consistently been negatively associated with mathematics performance (Ashcraft, Krause, & Hopko, 2007; Eden, Heine, & Jacobs, 2013; Hembree, 1990). The nature of the association between mathematical anxiety and performance remains unclear. Some studies suggest that poor performance and failure lead to mathematical anxiety (Ashcraft et al., 2007; Cipora, Szczygiel, Willmes, & Nuerk, 2015; Ma & Xu, 2004; Maloney, Risko, Ansari, & Fugelsang, 2010), which is particularly evident in adolescence and adulthood (Jameson, 2013). The relationship may also be moderated by motivation. A linear relationship between mathematical anxiety and performance has been found in individuals with low maths motivation. A curvilinear relationship was observed in highly motivated individuals: Low and high levels of maths anxiety were associated with poor performance and intermediate levels of maths anxiety with optimal mathematics performance (Wang et al., 2014).

The direction of causation can also be reversed, so that greater mathematical anxiety contributes to lower mathematics performance (e.g., Chinn, 2009; and also see reviews of Carey, Hill, Devine, & Szucs, 2016; Dowker, Sarkar, & Looi, 2016). This could be due to a direct effect of mathematical anxiety on mathematics-related cognitive abilities, such as working memory (see reviews of Carey et al., 2016; Dowker et al., 2016). Moreover, some evidence points to the bidirectional relationship between mathematical anxiety and mathematics performance, whereby mathematical anxiety and mathematics performance influence each other (Carey et al., 2016).

This study aimed to investigate:

(1) Whether average and variance differences exist between the UK, Russian, and Chinese 6- to 9-year-old children in mathematically relevant traits.

Previous studies have investigated cross-cultural differences in a number of mathematical abilities, including counting (e.g., Fuson & Kwon, 1992; Mark & Dowker, 2015;
Miller & Stigler, 1987; Song & Ginsburg, 1988; Xenidou-Dervou, Gilmore, van der Schoot, & van Lieshout, 2015); numerical characteristics (e.g., Göbel, Maier, & Shaki, 2015); place-value understanding (e.g., Mark & Dowker, 2015; Miller & Stigler, 1987; Miura, Kim, Chang, & Okamoto, 1988); number line estimation (e.g., Helmreich et al., 2011; Siegler & Mu, 2008); digit span (e.g., Chen, Cowell, Varley, & Wang, 2009; Yang et al., 2012); approximate number sense (e.g., Rodic et al., 2015); simple and complex arithmetic (e.g., Dowker, Bala, & Lloyd, 2008; Gatobu, Arocha, & Hoffman-Goetz, 2014; Geary et al., 1993; Laski & Yu, 2014; Robinson & Beatch, 2016; Rodic et al., 2015; Shen, Vasilyeva, & Laski, 2016; Stevenson et al., 1990; Vasilyeva, Laski, & Shen, 2015; Xenidou-Dervou et al., 2015; Zhou, Peverly, & Lin, 2005); word problems (e.g., Stevenson et al., 1990); overall mathematical ability, including PISA and TIMSS assessments (e.g., Caro, Lenkeit, & Kyriakides, 2016; He, Buchholz, & Klieme, 2017; Jak, 2017; Lowrie, Logan, & Ramful, 2016; Min, Cortina, & Miller, 2016; Pitchford & Outhwaite, 2016; Ryoo et al., 2014; Schachner, He, Heizmann, & Van de Vijver, 2017; Stevenson, Lee, & Stigler, 1986; Weis, Trommsdorff, & Muñoz, 2016; Zhou et al., 2005); and maths anxiety (e.g., Ahn, Usher, Butz, & Bong, 2016).

However, previous research is not sufficient in the following three areas: First, differences in symbolic magnitude comparison and maths reasoning have not been investigated, with the exception of one study that focused on differences in two-digit number comparison of Welsh and English children (Dowker et al., 2008). The advantage of Welsh children in two-digit number comparison found in the study might stem from irregular number system of English language.

Second, the majority of previous studies focused on cross-cultural differences between children from East Asia and the United States, between children from East Asia and Western Europe, or children from two countries in Western Europe. Very few studies examined differences in children’s ability across Russia, China, and Western Europe (Rodic et al., 2015; Shen et al., 2016). These comparisons are informative, considering the differences in cultural and educational practices across these populations.

Third, the cross-cultural differences in maths anxiety demonstrated in previous studies may have come from economic differences and consequent educational differences (e.g., the Philippines vs. the United States in Ahn et al., 2016). Cross-cultural research is needed to investigate potential differences in maths anxiety in similar economic circumstances.

(2) Whether cross-cultural differences exist in the interrelationships between arithmetic skills and mathematically relevant traits in 6- to 9-year-old children.

Several studies have found a longitudinal link between non-symbolic magnitude estimation and mathematical skills (Halberda & Feigenson, 2008; Mazzocco, Feigenson, & Halberda, 2011), while other studies have not found this link (e.g., De Smedt & Gilmore, 2011; Holloway & Ansari, 2009; Rodic et al., 2015; Soltész, Szücs, & Szücs, 2010). Further research is needed to evaluate whether the inconsistencies in the literature result from publication bias, when evidence for the absence of the associations is not reported; research limitations, such as small sample sizes; and differences in sample characteristics, such as age or cultural background of participants. Differences in measures may also contribute to the observed inconsistencies. For example, it has been found that non-symbolic magnitude processing is related to arithmetic fluency rather than mathematical problem-solving (Cui, Zhang, Cheng, Li, & Zhou, 2017; Zhang et al., 2016; Zhou, Wei, Zhang, Cui, & Chen, 2015). In addition, as described above, previous studies mostly focused on cross-cultural differences in different aspects of mathematical ability, rather than on interrelationships between arithmetic skills and mathematically relevant cognitive abilities. One study has investigated the interrelationships between arithmetic
skills and symbolic and non-symbolic numerical processing in 5- to 7-year-old preschool children from the United Kingdom, Russia, China, and Kyrgyzstan. The study found that arithmetic skills relied on symbolic numerical processing but not non-symbolic numerical processing (Rodic et al., 2015). The relationship between arithmetic and symbolic numerical processing was also found in third- to sixth-grade pupils in China (Wei, Lu, et al., 2012; Zhang et al., 2016). Several previous studies investigated cross-cultural differences in relationships between mathematics achievement and motivation, such as maths self-concept and maths self-efficacy (e.g., Marsh, 2016). However, little research is available on the cross-cultural differences in the links between maths ability and maths anxiety. This study investigates the interrelationships between arithmetic skills, numerical processing (both symbolic and non-symbolic), mathematical reasoning, and maths anxiety.

The study tests two main hypotheses:

(1) There are cross-cultural differences in mathematics performance among the UK, Russian, and Chinese 6- to 9-year-old children. Specifically, Chinese children will outperform UK and Russian peers in symbolic numerical comparison and mathematical reasoning. The expected differences might stem from differences in regularity of linguistic number system, as well as advantage of Chinese children in counting abilities and concept understanding, such as place-value understanding and number line estimation.

(2) Interrelationships between arithmetic skills and mathematically relevant traits are similar for children in primary schools of the three countries, consistent with previous findings (Rodic et al., 2015; Wei, Lu, et al., 2012; Zhang et al., 2016).

Methods

Participants

The participants were recruited from primary schools in the United Kingdom, Russia, and China. Children in Russia start their primary education a year later (at 7 years of age) than children in the United Kingdom and China. To match the groups on chronological age, Russian participants were in the second year of primary education, whereas the UK and Chinese participants were in the third year.

The sample consisted of 890 6- to 9-year-old children from the United Kingdom, Russia, and China. There were 73 UK participants (34 boys; mean age = 97.56 months, range 80–105 months); 421 Russian participants (232 boys; mean age = 105.41, range = 88–112 months); and 396 Chinese participants (221 boys; mean age = 104.27, range = 95–110 months). The children in the UK sample were recruited from five State schools in London Greater area. The children in the Russian sample were recruited from 15 State schools. The children in the Chinese sample came from 12 State schools. None of the schools operated any special intake selection.

Measures and procedure

A total of eight tests were computerized using Web-based applications in the ‘Online Psychological Experimental System (OPES)’ (www.dweipsy.com/lattice) (e.g., Cui, Georgiou, et al., 2017; Cui, Zhang, et al., 2017; Rodic et al., 2015; Zhang et al., 2016; Zhou et al., 2015). Children completed an online battery of tests in their schools, supervised by a researcher. The tasks (see Figure 1) were administered in the following order: Mental rotation, Choice reaction time, Non-symbolic comparison of numerosity,
Symbolic number magnitude comparison, Simple subtraction, Number series completion, Raven’s progressive matrices, and Mathematical anxiety questionnaire. Children were offered time to rest between the last two tasks. Mathematical anxiety questionnaire was completed last and followed general cognitive ability (rather than numerical) test, to minimize potential effects of prior tests on mathematical anxiety.

Children responded by pressing keys ‘Q’ or ‘P’ (and corresponding computer keys in Russian and Chinese), which were marked with colourful stickers. Responses on the Mathematical anxiety questionnaire were recorded by a researcher. Accuracy (ACC) and RT (in milliseconds) were recorded for Choice reaction time, Non-symbolic comparison of numerosity, and Symbolic number magnitude comparison tasks. For the Mathematical anxiety questionnaire, the total score was calculated by adding up the responses on a 5-point Likert scale for the whole questionnaire. For the rest of the tasks (i.e., Mental rotation, Simple subtraction, Number series completion, and Raven’s progressive matrices), the score was calculated by subtracting incorrect from correct responses, to correct for guessing. The tasks were grouped in the following categories: (1) general skills and IQ; (2) spatial ability; (3) symbolic number understanding; (4) non-symbolic estimation; (5) operating with numbers (arithmetic) and numerical reasoning; and (6) mathematical anxiety.

The study and consent procedure were approved by the Ethics Committee of Goldsmiths, University of London.

**General skills and IQ**

The Choice reaction time task, adapted from the Dyscalculia screener (Butterworth, 2003), was used to measure the processing speed. A dot appeared at the 30 degree angle on the left (15 trials) or right (15 trials) side of the fixation ‘+’, with interstimulus interval between 1,500 and 3,000 ms. Children’s accuracy (pressing the relevant button) and speed of responses were recorded. This task can be considered as a ‘baseline task’, controlling for influence of individual differences in processing speed (time associated

![Figure 1. Illustration of tasks used in the study, in the order of presentation.](image)
with pressing buttons), allowing for evaluation of pure processing time for other cognitive and mathematical processing (e.g., Butterworth, 2003, see page 13, using a simple reaction time task as baseline).

The Raven’s progressive matrices task was adapted from the legal copy of Raven’s Progressive Matrices and computerized on the OPES (Raven, Raven, & Court, 1998). The task measures abstract reasoning and serves as an index of \( g \). Children were presented with an incomplete figure with six segments underneath it. The child had to identify the correct segment to complete the figure’s intrinsically regular pattern. Children had to go through as many trials as they could in 4 min. The test was shortened to 80 items, including 44 items from the Standard Progressive Matrices (12 items from the first set and eight items from each of the other four sets) and 36 items from the Advanced Progressive Matrices. The shortened version has been used in previous studies to test general IQ or abstract reasoning (e.g., Bors & Vigneau, 2003; Bouma, Mulder, & Lindeboom, 1996; Vigneau & Bors, 2001; Vigneau, Caissie, & Bors, 2006; Wang, Sun, & Zhou, 2016; Wei, Lu, \textit{et al.}, 2012; Zhang \textit{et al.}, 2016; Zhou \textit{et al.}, 2015). The short time limit of 4 min was based on previous research with children that demonstrated that having longer time did not result in greater proportion of correct responses. For example, in one study, the number of unanswered items was unrelated to APM (Raven’s Advanced Progressive Matrices) score, with improvements in performance not based on a strategy to respond to more items (Bors & Vigneau, 2003). This adaptation of the Raven Progressive Matrices has acceptable split-half reliability according to previous studies (.84 – see Cui, Zhang, \textit{et al.}, 2017; and .83 – see Wei, Yuan, \textit{et al.}, 2012).

Adjusted number of correct trials was used to control for the effect of guessing in multiple choice tests (e.g., Cirino, 2011; Hedden & Yoon, 2006; Salthouse, 1994; Salthouse & Meinz, 1995). The score was calculated by subtracting the number of incorrect responses from the number of correct responses, following the Guilford correction formula ‘\( S = R-W/(n-1) \)’ (\( S \): the adjusted number of items that the participants can actually perform without the aid of chance; \( R \): the number of correct responses; \( W \): the number of wrong responses; and \( n \): the number of alternative responses to each item) (Guilford & Guilford, 1936). This correction procedure has been utilized recently in studies of mathematical cognition (Cirino, 2011; Cui, Zhang, \textit{et al.}, 2017; Zhou \textit{et al.}, 2015) and cognition in general (Hedden & Yoon, 2006; Putz, Gaulin, Sporter, & McBurney, 2004; Salthouse, 1994). In this study, the number of alternative answers is 2; therefore, \((n-1) = 1\). Consequently, the scores ranged from \(-80\) to 80.

**Spatial ability**

The Mental rotation task, adapted from Shepard and Metzler (1971), was used to assess children’s ability to mentally rotate three-dimensional images. One (target) image was presented at the top of the screen with two potential answers on the left and right bottom parts of the screen. Children had to select the image that was matching the top one. The matching images were rotated from \(15^\circ\) to \(345^\circ\). Children had to go through as many trials (max 180) as they could in 3 min.

**Symbolic number understanding**

The Symbolic number magnitude comparison task, adapted from Girelli, Lucangeli, and Butterworth (2000), used a Stroop-like paradigm. The task assessed the ability to compare numerical values of numbers that varied in physical size (1:2 size ratio). Children had
5 seconds to judge which of the two single-digit numbers, appearing simultaneously on
the screen, was larger in numerical magnitude, ignoring the differences in physical size.
There were three types of trials: neutral (both digits were of the same physical size),
congruent (a numerically larger digit was also physically larger), and incongruent
(numerically smaller digit was physically larger). There were three sessions of 28 trials,
separated by 10-second resting periods.

**Non-symbolic estimation**
The Non-symbolic comparison of numerosity, adapted from Baroody and Ginsburg
(1990), was used to measure non-symbolic number sense. Two sets of dots of varying sizes
(5–12 dots; ratios 2:3, 5:7, and 3:4) were presented simultaneously on the screen.
Children were asked to estimate (without counting) which of the two sets contained more
dots. There were 36 trials in total. As proposed by Gebuis and Reynvoet (2011, 2012), the
following five types of visual stimulus parameters were controlled in the test: total surface
area, envelope area or convex hull, item size, density (envelope area divided by total
surface), and circumference. Even after considering the five visual properties, the
accuracy across all trials of this test was still ratio-dependent, \( r (113) = 0.31, p < 0.001 \)
(see Zhou *et al.*, 2015).

**Operating with numbers**
The Simple subtraction task, developed according to the theory in previous work
(Landerl, Bevan, & Butterworth, 2004; Girelli *et al.*, 2000), assessed early arithmetic
ability. The problem was presented in the middle of the screen with two candidate
answers beneath it. Children had to choose the correct answer in as many problems as
they could (max 92) in 2 min. All minuends were smaller than 18, and all differences were
single-digit numbers. Correct and incorrect answers were within the range of each other
plus or minus 3.

The Number series completion task, adapted from Smith, Fernandes, and Strand
(2001), measured logical numerical reasoning. Children completed the sequence of
numbers presented on the screen (e.g., 2, 4, 6, 8), by choosing one of the two candidate
answers (e.g., 9 or 10) presented below it. Children had to go through as many trials (max
43) as they could in 4 min.

**Maths anxiety**
The Revised Mathematics Anxiety Rating Scale (RMARS; Alexander & Martray, 1989)
included 25 items, such as ‘How would you feel having math as part of a test’. Each
statement was read out, and children reported (on the scale of 1 = not at all to 5 = very
strongly) how anxious they would feel in these situations. The full list of items in English is
presented in Table S1, together with information on adaptation for administration in
Russian and Chinese.

**Results**
Descriptive statistics on raw data for all tasks can be seen in Table S2. The data were age-
regressed for further analyses to eliminate any effects of chronological age. Table S3
presents the results of the sex differences analyses, showing non-significant or negligible effects of sex.

**Cross-cultural comparisons**

Table S4 presents the results of ANOVAs and Bonferroni *post-hoc* analyses conducted for all measures on uncorrected scores. All tests exhibited significant cross-cultural differences. UK children on average spent more time and achieved higher accuracy, while Chinese children tended to use less time leading to a lower accuracy in non-time-limited tests (choice reaction time, symbolic magnitude comparison, and non-symbolic comparison of numerosity).

We have calculated the speed–accuracy trade-off effect for the symbolic and non-symbolic magnitude comparison tasks, using an inverse efficiency measure – IES (Townsend & Ashby, 1978; e.g., Pouw, Van, Zwaan, & Paas, 2016; Setti, Borghi, & Tessari, 2009). IES is estimated as reaction time divided by proportion correct responses, with higher scores indicating using more time to obtain a higher proportion of correct responses. The IES scores for the UK, Russian, and Chinese samples were 11.96, 10.85, and 8.66 for symbolic magnitude comparison; and 12.82, 12.44, and 10.63 for non-symbolic numerosity comparison, respectively.

**Figure 2.** Mean accuracy, mean RTs, and mean correct–incorrect responses with standard error bars (1SD) for the tasks where significant differences emerged between samples: (1) Chinese > UK, Russian; (1a) Russian > Chinese; (2) Chinese > UK, Russian; (2a) UK > Chinese, and Russian > Chinese; (3) Chinese > UK, Russian; (3a) UK > Russian, Chinese, and Russian > Chinese; (4) Chinese > UK, Russian; (5) Chinese > UK, Russian, and UK > Russian; (6) Chinese, Russian > UK; and (7) Russian > UK, Chinese.

Note: All data were age- and IQ-regressed and cleaned for outliers (± 3SD).
ANOVAs and Bonferroni post-hoc analyses were conducted for all measures on scores corrected for raw scores of Raven’s progressive matrices (to control for chronological age and grade differences). The majority of tasks showed significant group differences in accuracy and/or RT (see Figure 2), with modest effect sizes, except performance on Mental rotation (spatial ability) and Mathematical anxiety.

There were significant differences among groups (all $p < .001$) in Choice reaction time, Symbolic number magnitude comparison, and Non-symbolic comparison of numerosity. Chinese children had faster reaction time and slightly lower accuracy than Russian/UK children (all $p < .05$).

Significant differences were also found among all groups (all $p < .001$) on Simple subtraction and Number series completion among groups, with Chinese children outperforming the UK and Russian children. That is, with limited time available for the two tasks, Chinese children obtained significantly higher adjusted correct scores (correct minus incorrect trials) than the UK and Russian children. The advantage of Chinese children in these tasks was present for scores both uncorrected (Table S4) and corrected for Raven’s progressive matrices (Figure 2) and Mental rotation (Table S5). This suggests that this advantage cannot be explained by differences in $g$ or spatial ability across the samples.

The two tasks that did not lead to differences across samples showed normal distributions. For Mental rotation, the values of skewness and kurtosis were as follows: for all samples combined, $- .362$ and $ .329$, respectively; for the UK sample, $ .008$ and $- .338$; for the Russian sample, $- .376$ and $-.162$; and for the Chinese sample, $ .193$ and $- .091$. For mathematical anxiety task, skewness and kurtosis were as follows: for all samples combined, $ .191$ and $-.497$; for the UK sample, $ .185$ and $-.751$; for the Russian sample, $ .258$ and $-.484$; and for the Chinese sample, $ .103$ and $-.696$.

ANOVAs and Bonferroni post-hoc analyses were also conducted for all measures on scores corrected for spatial ability, as measured by Mental rotation. Most of the tasks showed significant group differences in accuracy and/or RT (see Table S5). The effect of Mental rotation on other measures is relatively small and did not change the significance of cross-cultural differences observed in the raw (uncorrected) data.

Regression analyses

A three-step hierarchical multiple regression was conducted separately in each sample with Simple subtraction as the criterion. Table 1 summarizes the variables included in the three steps and the results of the regression.

All three samples exhibited a significant role of general cognitive processing (Choice reaction time, Mental rotation, and Raven’s progressive matrices), basic numerical processing (Symbolic number magnitude comparison and Non-symbolic comparison of numerosity), Number series reasoning, and gender in variation of Simple subtraction (except only a marginal significance of general cognitive processing in the UK sample).

Next, we grouped all samples together and repeated the same analysis, with Sample as an additional predictor. Together, RT and ACC of the Choice reaction time task, and correct - incorrect responses for Mental rotation and Raven’s progressive matrices explained 25.2% of the variance in Simple subtraction. RT and ACC of the Symbolic number magnitude comparison task and the Non-symbolic comparison of numerosity task explained additional 13.3% of the variance. Number series completion (correct - incorrect responses) and Gender explained 11.4% of the variance. Finally, Sample explained additional 8.5% of the variance in Simple subtraction.
Table 1. Significant predictors of hierarchical multiple linear regression

<table>
<thead>
<tr>
<th>Sample</th>
<th>Significant predictors</th>
<th>B</th>
<th>SE b</th>
<th>95% confidence interval</th>
<th>F</th>
<th>R² at each step</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 65</td>
<td>Choice reaction time (RT)</td>
<td>−0.004</td>
<td>0.009</td>
<td>−0.022</td>
<td>0.013</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Choice reaction time (acc.)</td>
<td>6.026</td>
<td>36.267</td>
<td>−12.214</td>
<td>132.736</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mental rotation</td>
<td>0.080</td>
<td>0.109</td>
<td>−0.138</td>
<td>0.298</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Raven’s</td>
<td>0.076</td>
<td>0.299</td>
<td>−0.522</td>
<td>0.673</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Symb. magnitude (RT)</td>
<td>−0.013*</td>
<td>0.005</td>
<td>−0.023</td>
<td>−0.003</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Symb. magnitude (acc.)</td>
<td>46.691</td>
<td>24.461</td>
<td>−2.255</td>
<td>95.638</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-symb. comparison (RT)</td>
<td>0.006</td>
<td>0.005</td>
<td>−0.004</td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-symb. comparison (acc.)</td>
<td>31.861*</td>
<td>13.857</td>
<td>4.133</td>
<td>59.588</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number series</td>
<td>0.659***</td>
<td>0.195</td>
<td>0.27</td>
<td>1.049</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>0.299</td>
<td>2.054</td>
<td>−3.814</td>
<td>4.412</td>
<td></td>
</tr>
<tr>
<td>Russian</td>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 288</td>
<td>Choice reaction time (RT)</td>
<td>−0.006</td>
<td>0.003</td>
<td>−0.012</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Choice reaction time (acc.)</td>
<td>16.533</td>
<td>8.694</td>
<td>−0.581</td>
<td>33.647</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mental rotation</td>
<td>0.128**</td>
<td>0.048</td>
<td>0.034</td>
<td>0.222</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Raven’s</td>
<td>0.566**</td>
<td>0.137</td>
<td>0.296</td>
<td>0.837</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Symb. magnitude (RT)</td>
<td>−0.009**</td>
<td>0.003</td>
<td>−0.014</td>
<td>−0.003</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Symb. magnitude (acc.)</td>
<td>37.546***</td>
<td>6.993</td>
<td>23.782</td>
<td>51.311</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-symb. comparison (RT)</td>
<td>−0.001</td>
<td>0.003</td>
<td>−0.006</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-symb. comparison (acc.)</td>
<td>6.824</td>
<td>5.089</td>
<td>−3.194</td>
<td>16.841</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number series</td>
<td>0.303***</td>
<td>0.074</td>
<td>0.157</td>
<td>0.448</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>−1.540</td>
<td>1.010</td>
<td>−3.529</td>
<td>0.448</td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Sample</th>
<th>Significant predictors</th>
<th>( B )</th>
<th>( SE , b )</th>
<th>95% confidence interval</th>
<th>( F )</th>
<th>( R^2 ) at each step</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( N = 396 )</td>
<td>Choice reaction time (RT)</td>
<td>-0.009***</td>
<td>0.003</td>
<td>-0.016</td>
<td>-0.003</td>
<td>21.627</td>
</tr>
<tr>
<td></td>
<td>Choice reaction time (acc.)</td>
<td>4.398</td>
<td>4.530</td>
<td>-4.508</td>
<td>13.304</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mental rotation</td>
<td>0.243***</td>
<td>0.055</td>
<td>0.136</td>
<td>0.351</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Raven's</td>
<td>0.357***</td>
<td>0.060</td>
<td>0.240</td>
<td>0.474</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Symb. magnitude (RT)</td>
<td>-0.008*</td>
<td>0.003</td>
<td>-0.014</td>
<td>-0.002</td>
<td>18.522</td>
</tr>
<tr>
<td></td>
<td>Symb. magnitude (acc.)</td>
<td>28.29***</td>
<td>6.855</td>
<td>14.815</td>
<td>41.765</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-symb. comparison (RT)</td>
<td>0.000</td>
<td>0.002</td>
<td>-0.004</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-symb. comparison (acc.)</td>
<td>24.322***</td>
<td>4.325</td>
<td>15.819</td>
<td>32.825</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Step 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number series</td>
<td>0.464***</td>
<td>0.088</td>
<td>0.291</td>
<td>0.636</td>
<td>14.588</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>0.622</td>
<td>0.704</td>
<td>-0.761</td>
<td>2.005</td>
<td></td>
</tr>
<tr>
<td>All samples</td>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>grouped</td>
<td>Choice reaction time (RT)</td>
<td>-0.018***</td>
<td>0.003</td>
<td>-0.023</td>
<td>-0.013</td>
<td>64.494</td>
</tr>
<tr>
<td>together</td>
<td>Choice reaction time (acc.)</td>
<td>0.155*</td>
<td>4.918</td>
<td>-9.499</td>
<td>9.810</td>
<td></td>
</tr>
<tr>
<td>( N = 749 )</td>
<td>Mental rotation</td>
<td>0.181***</td>
<td>0.040</td>
<td>0.102</td>
<td>0.260</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Raven's</td>
<td>0.674***</td>
<td>0.064</td>
<td>0.548</td>
<td>0.800</td>
<td></td>
</tr>
</tbody>
</table>

Continued
Table 1. (Continued)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Significant predictors</th>
<th>B</th>
<th>SE b</th>
<th>95% confidence interval</th>
<th>F</th>
<th>R² at each step</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.133***</td>
</tr>
<tr>
<td></td>
<td>Symb. magnitude (RT)</td>
<td>-0.022***</td>
<td>0.002</td>
<td>-0.026</td>
<td>41.151</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Symb. magnitude (acc.)</td>
<td>38.076***</td>
<td>5.54</td>
<td>27.226</td>
<td>48.927</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-symb. comparison (RT)</td>
<td>7.889*</td>
<td>3.65</td>
<td>0.714</td>
<td>0.018</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-symb. comparison (acc.)</td>
<td>0.003</td>
<td>0.002</td>
<td>0.000</td>
<td>0.018</td>
<td></td>
</tr>
<tr>
<td>Step 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.114***</td>
</tr>
<tr>
<td></td>
<td>Number series</td>
<td>0.670***</td>
<td>0.051</td>
<td>0.569</td>
<td>0.771</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>-0.411</td>
<td>0.611</td>
<td>-1.610</td>
<td>0.788</td>
<td></td>
</tr>
<tr>
<td>Step 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.085***</td>
</tr>
<tr>
<td></td>
<td>Sample</td>
<td>6.813***</td>
<td>0.547</td>
<td>5.739</td>
<td>7.887</td>
<td></td>
</tr>
</tbody>
</table>

Notes. The Choice reaction time (RT and accuracy), Mental rotation (correct - incorrect responses), Raven’s (correct - incorrect responses) were entered at step 1; RT and accuracy of both Symbolic number magnitude comparison and Non-symbolic comparison of numerosity were entered at step 2, Number series (correct - incorrect responses) and gender were entered at step 3, and Sample (i.e., nation) was entered at step 4; for Magnitude comparison task the trials were collapsed together as no congruency effects were found (high performance on average in all trials); for Non-symbolic comparison of numerosity, the trials were collapsed together as no ratio effects were found (high performance on average in all trials); the results for congruency and ratio analyses are available from the corresponding author. The bold values indicate significant change in $R^2$ in each step.

* $p < .05$; ** $p < .01$; *** $p < .001$. Only participants with complete data for all variables were included in this analysis.
Discussion

The first aim of this study was to investigate average and variance differences between the UK, Russian, and Chinese early primary school children in arithmetic, related cognitive skills, and mathematical anxiety.

Sample explained 8.5% of the variance of the speed in which the children performed the baseline task (i.e., Choice reaction time task, which eliminates influence of individual differences in processing speed to see the pure processing time of other general cognitive processing and mathematical processing rather than the time to press buttons), with Chinese children showing faster performance than the UK and Russian children. Accuracy on this task, which was approaching ceiling (~95% for all children), produced small (1.8%) differences.

The measure that required good understanding of the numerical value of numerals and the capacity to order numerosities by their size also showed small (4.3%) differences in accuracy. The accuracy on this task was high overall, with all children scoring in the range of 90%. For the RT measure of this task, 21% of the variance was explained by Sample, indicating significantly faster performance by the Chinese children over the other two samples.

For the non-symbolic magnitude processing task, for both accuracy and RT, ~6% of the variance was explained by the Sample. Chinese children performed faster than the remaining two samples. Once again, the accuracy was high (~90%). Chinese children performed significantly less accurate than both the UK and Russian children according to our results of speed–accuracy trade-off effect.

Overall, the RT measure seems to better discriminate children’s performance at this stage (see Butterworth, 2003). Processing speed is an important measure, particularly for basic overlearned mathematical abilities, and has previously been found to predict mathematics performance in 7-year-old children (Bull & Johnston, 1997). Some studies also suggest that slow reaction time on tasks, such as number identification, visual number matching, magnitude comparison, and digit encoding, can be used to identify children with mathematical difficulties (Geary et al., 1993).

For Number series completion, a numerical reasoning task, 17.3% of the variance was explained by the Sample, with Chinese children showing the best performance.

For the Simple subtraction task, 28% of the variance was explained by the Sample. In line with the previous research with both preschool (Rodic et al., 2015) and older children (Imbo & LeFevre, 2009; Mullis et al., 2008; OECD, 2010), Chinese 6- to 9-year-olds outperformed both the UK and Russian samples.

The Mental rotation task did not show any cross-cultural differences in this study. It is possible that visuospatial advantage of Asian adults (Sakamoto & Spiers, 2014), attributed to a potential impact of spatially complex character-based writing system on the development of spatial ability (Flaherty & Connolly, 1995), is present only in more advanced users of character-based writing systems. Chinese children in our sample might not have been exposed to the spatial complexity of the character-based writing system long enough to lead to the advantage on spatial ability tests.

The levels of Mathematical anxiety did not differ significantly between the samples. It is also possible that mathematical anxiety is not yet pronounced at this age (Jameson, 2013). However, several previous studies, using different measures, have shown that MA was present in younger primary students (Harari, Vukovic, & Bailey, 2013; Krinzinger, Kaufmann, & Willmes, 2009; Ramirez, Gunderson, Levine, & Beilock, 2013). It is possible that the self-report measure used in this study is not precise in measuring individual differences in mathematical anxiety in young children. In particular, some children might
not have been familiar with some of the items. The overall variability in the scores was quite small, and the average level of anxiety was quite low. The relatively low MA scores might be related to higher scores in subtraction task in the current study compared to peers (e.g., lower scores were found in Zhang et al., 2016 and Zhou et al., 2015, using the same subtraction task). Better performance may be associated with lower anxiety through unilateral or reciprocal causal links (e.g., better performance leading to greater confidence). Alternatively, the link may reflect other factors, such as covering simpler material in previous lessons or teachers’ characteristics.

In line with previous research with participants of different ages (Geary et al., 1993; Leung, 2006; Rodic et al., 2015; Song & Ginsburg, 1988), Chinese 6- to 9-year-old children outperformed the other two samples on both simple arithmetic and mathematical reasoning tasks, with moderate effect sizes (17–28%). In comparison with the Rodic et al. (2015) study, where Chinese preschool children showed advantage with 13% effect size, the current study indicates the increase of this effect (28%), suggesting that the advantage might be increasing with age and/or formal instruction. This finding suggests that in addition to linguistic advantage of Chinese children (Dehaene, 1997) and extra time Chinese parents spend teaching and practising with their children before they reach school (Huntsinger, Jose, Larson, Balsink Krieg, & Shaligram, 2000), the Chinese formal educational system might provide additional advantage at the early primary school ages. For example, some research suggests that parents in China are better informed by the schools of what is expected of their children and are thus able to provide more adequate help (Miller et al., 2005). In addition, mothers of first graders in China show increased involvement as their children start formal education, in comparison with US mothers (Chao, 2001).

The observed overall faster performance of Chinese children on measures of symbolic and non-symbolic magnitude comparison may indicate more in-depth knowledge of magnitudes and Arabic numerals, which in turn could also lead to better performance in arithmetic problem-solving.

Overall, the results suggest that culture has an effect on performance. For example, the results of the multiple regression on all three samples combined showed that Sample explained 8.5% of the variance in Simple subtraction, additional to the effects of Raven’s progressive matrices, Mental rotation, and other abilities. The actual effects of culture are likely to be even greater as some effects may be ‘removed’ when controlling for other abilities on which samples also differed.

The second aim of this study was to investigate whether cognitive skills that predict arithmetic in early primary school differ between the three populations. Understanding numbers and magnitudes, as well as mathematical reasoning, predicted arithmetic in all three samples. This is in line with the previous research which shows that the ability to process numerical magnitudes of symbolic numbers is important for the development of arithmetic skills and is positively related to mathematics achievement (Booth & Siegler, 2006, 2008; Castronovo & Göbel, 2012; De Smedt et al., 2009; Gilmore et al., 2010; Holloway & Ansari, 2009; Jordan, Kaplan, Ramineni, & Locuniak, 2009). In line with previous studies, Non-symbolic magnitude comparison task predicted arithmetic in the UK and Chinese samples (Libertus, Feigenson, & Halberda, 2011; Mazzocco et al., 2011). It is possible that shorter period of instruction in the Russian sample (second vs third grade) led to an absence of this association.

In line with previous research (Gottfredson & Deary, 2004; Rohde & Thompson, 2007; Strenberg, Grigorenko, & Bundy, 2001), Mental rotation and Raven’s progressive matrices predicted arithmetic in the Russian and Chinese samples. The absence of the effect in the UK sample could be due to a reduced power of a small sample size.
Conclusion and future directions
In line with previous research, Chinese children outperformed the other samples on early arithmetic and early mathematical reasoning. No significant differences in maths anxiety have been found. Despite the cross-cultural differences in performance, similar mechanisms of mathematical development seem to be operating in all three populations.

The cross-sample comparisons should be interpreted in the light of potential sample differences (age, grade, and variance differences). Further longitudinal research, with increased sample sizes, additional populations, and assessed cultural/educational features, is needed to clarify the sources of the observed differences.

Acknowledgements
The authors gratefully acknowledge the support from the British Academy, the Royal Society, the Open Research Fund of the State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University; and the Grant No. 16-36-01102 from the Russian Foundation for Basic Research.

References


Received 29 September 2016; revised version received 14 April 2018

**Supporting Information**

Additional supplemental material may be found online in the Supporting Information section at the end of the article:

**Table S1.** The Revised Mathematics Anxiety Rating Scale (RMARS) in English.

**Table S2.** Descriptive statistics on raw scores for all tasks - UK, Russian & Chinese samples.

**Table S3.** Sex differences for all tasks per country.

**Table S4.** ANOVA results on raw (uncorrected) scores.

**Table S5.** ANOVA results on scores corrected for Mental rotation.