NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Yulia Kuzmina, Elen Yusupova

THE NUMERICAL RATIO EFFECT FOR DIGITS AND NUMBER WORDS: THE LACK OF INDIVIDUAL DIFFERENCES

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: PSYCHOLOGY

WP BRP 135/PSY/2023

Yulia Kuzmina[1] *, Elen Yusupova*[2]

THE NUMERICAL RATIO EFFECT FOR DIGITS AND NUMBER WORDS: THE LACK OF INDIVIDUAL DIFFERENCES[3]

In the current study, we inspected the numerical ratio effect (NRE) for digits and number words in a comparison task using a large sample of third graders (N=1383, mean age was 9.84 years, 49% were girls) and applied linear mixed effects models to estimate the average NRE and its between-individual variability. An analysis demonstrated that the sample average NRE was significant for both digits and number words. For digits and number words, reaction time (RT) increased, while the numerical ratio between two compared numbers increased, although the patterns of changes were different for digits and number words. An inspection of between-individual variance in the NRE revealed that the NRE has no between-individual variance independent of variance in RT.

JEL Classification: Z

Keywords: numerical ratio effect, symbolic representation, digits, number words, individual differences

[1] Centre for Psychometric and Measurement in Education, Institute of Education, HSE University, jkuzmina@hse.ru
[2] Centre for Psychometric and Measurement in Education, Institute of Education, HSE University, eabdurakhmanova@hse.ru

[3] This Working Paper is an output of a research project implemented within NRU HSE's Annual Thematic Plan for Basic and Applied Research. Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.

Introduction

Humans possess the ability to process quantitative information in a symbolic format with digits and number words. The way in which these numerical symbols are organized is referred to as the external numerosity system (Zhang & Norman, 1995). The processing of the external numerosity system occurs by creating an internal numerosity representation, i.e., a mental "map" of external numerosity. To describe the internal numerosity representation, many scholars have used the metaphor of the "mental number line," along which numerosities are located (e.g., Izard & Dehaene, 2008; Nuerk et al., 2004; Nuerk et al., 2011;). The precision of the symbolic numerosity representation reflects the degree of accuracy and speed of establishment of correspondence between external and internal numerosity systems.

The precision of symbolic representation is often measured by a digit comparison test (e.g.,; De Smedt et al., 2013; Kolkman et al., 2013). Several indicators of precision can be used: accuracy (the proportion of correct answers), speed (e.g., the average reaction time for correct answers) and the numerical ratio (or distance) effect (NRE or NDE) (e.g., Bartelet et al., 2014; Holloway & Ansari, 2009; Rousselle & Noël, 2007). The NRE and NDE are manifested in larger RT and lower accuracy when comparing numerosities that are closer to each other on the mental number line and have a larger numerical ratio (or shorter distance) than when comparing numerosities with a smaller numerical ratio between them (e.g., Holloway & Ansari, 2009; Lyons, Nuerk, & Ansari, 2015; Maloney et al., 2010). For example, the comparison of 6 and 8 (numerical ratio is 0.75) requires more time than the comparison of 4 and 8 (numerical ratio is 0.50).

For many studies, the NRE (or NDE) is the core feature of internal numerosity representation, indicating the imprecision of the mental representation of numerosity wherein a lower NRE indicates a more precise representation (e.g., Holloway & Ansari, 2009; Mundy & Gilmore, 2009; Núñez-Peña & Suárez-Pellicioni, 2014). The NRE is supposed to arise due to the overlapping of Gaussian curves, reflecting the activation of neurons coding for numbers (e.g., Nieder & Miller, 2003). However, some authors have assumed that the NRE does not imply the overlapping of internal representations for two quantities (e.g., Van Opstal et al., 2008) and instead may arise from relative word frequency and response selection processes (Verguts & Van Opstal, 2005). Although the NRE and NDE are often used interchangeably, it is suggested that the NRE better reflects the properties of internal numerosity representation (Dietrich, Huber, & Nuerk, 2015).

Although the NRE in comparison tasks is detectable in most studies, some authors have questioned the use of the NRE as an indicator of numerosity representation precision (Lyons et al., 2015). In particular, a lower NRE value can indicate different phenomena. First, the NRE may be small or absent due to precise symbolic representation. Second, the NRE may be small or absent because the ratio between two numerosities does not affect the accuracy or RT of numerosity discrimination, and consequently, it may be an inappropriate indicator of symbolic representation precision. Third, it was suggested that a small NRE or its absence might indicate that the comparison of two digits is automatic and fast; thus, access to internal numerosity representation is not mandatory (e.g., Sasanguie & Reynvoet, 2014). Therefore, smaller NRE values do not necessarily imply more precise symbolic representation.

Many studies have emphasized that the NRE/NDE is format-independent and identified for symbolic (digits and number words) and nonsymbolic comparisons (e.g., Kadosh, Henik, & Rubinstein, 2008; Krajcsi et al., 2016; Lyons et al., 2015; Maloney et al., 2010). However, some studies have revealed differences in the NRE/NDE for digits, number words and nonsymbolic comparisons (e.g., Lyons et al., 2015; Maloney et al., 2010). In particular, an inspection of the NDE for digits and number words demonstrated that the NDE was smaller for number words than for digits (Kadosh et al., 2008).

The NRE/NDE for symbolic and nonsymbolic formats are primarily studied in experimental studies with a focus on the detection of average effects or common mechanisms (e.g., Borsboom et al., 2009). In these studies, between-individual differences in effects are treated as "noise" and are not taken into account. However, sometimes the NRE or NDE are used in correlational studies as indicators of the individual level of the precision of numerosity representation. Particularly, it was determined that a smaller NRE (NDE) for Arabic digits was associated with higher math achievement (e.g., Askenazi et al., 2009; De Smedt et al., 2009; Göbel et al., 2014; Lonneman et al., 2011). It has also been shown that NRE for digits and nonsymbolic comparisons is not correlated at the individual level (e.g., Lyons et al., 2015). The absence of correlation between NREs for different formats was discussed as the indicator that nonsymbolic and symbolic NRE have different sources (e.g., Lyons et al., 2015).

However, the lack of correlation between NREs for different formats might be due to low between-individual variance of these indicators (Borsboom et al., 2009). In turn, low variance of NRE can lead to low reliability of these measures (e.g., Ponterotto & Ruckdeschel, 2007; Henson, 2001). There is evidence that NREs for different formats have low reliability (e.g., Maloney et al., 2010; Sasanguie et al., 2011). In particular, the split-half and test-retest reliability of the NDE/NRE was low for nonsymbolic comparison (e.g., Dietrich et al., 2016; Sasanguie et al., 2011) and for digit comparison (Maloney et al., 2010).

In addition, the lack of convergent validity of the NDE/NRE for nonsymbolic and symbolic comparison tasks has been shown to be due to low correlations between the NDE/NRE and other measures, such as accuracy and average RT (Dietrich et al., 2016; Inglis & Gilmore, 2016; Price et al., 2012). It has also been found that a large number of pupils did not show statistically meaningful NRE (Lyons et al., 2015). Assuming the abovementioned issues with NRE, some scholars have cautioned against the use of the NRE/NDE for assessing individual differences in numerosity processing (e.g., Maloney et al., 2010; Lyons et al., 2015).

Previous studies have mostly focused on assessing, comparing and estimating the correlation between NREs for digits and nonsymbolic comparisons, while NRE for number words has been less studied from the perspective of between-individual differences. It is not well known whether between-individual differences in NRE for number words exist and how NRE for number words is associated with NRE for digits.

In the current study, we aim to estimate the significance of between-individual differences in NRE (for RT) in a comparison task for number words and digits. In previous studies of NRE, two strategies for calculating the NRE were used. The first strategy implies calculating the difference between RT in comparison numbers with a high ratio (close distance) and RT in comparison numbers with a low ratio (large distance) (e.g., Dehaene & Akhavein, 1995). The second strategy implies the estimation of regression models for each participant with RT as the dependent variable and the numerical ratio (NR) for each item as a predictor (e.g., Lyons et al., 2015). The coefficient of the NR variable indicates an individual NRE. To obtain the sample mean NRE, individual NREs are averaged. In contrast with previous studies, we used mixed effects models (also known as multilevel regression models), which enabled us to estimate the sample mean NRE (fixed effect) and between-individual variability in this effect (random effect for slope) in a large sample of third graders (N=1383). The multilevel regression approach has several advantages over linear regression, which is described in more detail in the Method section.

Method

Sample

The sample for this study consisted of third grade pupils from Russia. The students were engaged in a longitudinal study of math and reading progress in elementary school (e.g., Ivanova et al., 2018). The assessment of the number comparison test was performed at the end of grade 3. Overall, 1383 pupils were involved in the number comparison study, 49% of whom were girls, and the average age at the end of grade 3 was 9.84 years (SD=0.34, range 9–11 years).

All procedures performed in the study were in accordance with the ethical standards of psychological studies. Informed consent was obtained from the parents of all participants.

Instrument and procedure

Pupils performed digit and written number word comparison tasks on computers. Numbers were presented on the screen simultaneously, and the task was to select which number (the one on right or on the left) was larger by pressing the key “arrow left” if the number was larger on the left and “arrow right” if the number was larger on the right. Each screen with digits or number words was presented for 2 sec., after which the screen with numbers disappeared and the gray screen was demonstrated. The child was given a 5-second period to answer the question. If an answer was not given within this period, the trial was recorded as missed, and the next screen with notification to continue the test was presented: “Press any key to continue.” This notification was also shown after the answer was given following each task. After pressing any key, the next item was presented.

The numbers varied from 1 to 9. Digit comparison and number word comparison tasks were demonstrated in mixed random order. This order was the same for all participants. There were 24 trials with digit comparisons and 24 trials with number word comparisons. The same pairs of numbers were used for digit and number word comparisons. For example, in digit format, a participant should select the largest number from pair “8 5”, while in number word format, s/he compared “eight five”. Hence, the set of trials in digit and number word formats was not different in the ratio between numbers and size (the sum of two numbers). The ratio between the two compared numerosities (smaller numerosity divided by larger numerosity) varied from 0.14 to 0.89 for digits and number words (average ratio was 0.51). The size (sum of two numbers) varied from 5 to 17. In half of the trials in each format, the larger number was on the left position, and in the other half, the larger number was on the right position.

Statistical approach

Before data analysis, the inspection of RT was performed. The answers that were given faster than 5 msec were transformed to missing answers to exclude random answers (Baayen & Milin, 2010).

We applied mixed effect models (also known as multilevel linear models, MLMs). Mixed effects models have several advantages compared to linear regressions with respect to the analysis of experimental data and cognitive tests (e.g., Brauer & Curtin, 2017; Field & Wright, 2011;

Hoffman & Rovine, 2007). For example, mixed effect models allow researchers to take into account nonindependence in data that occurs when participants are executing the same tasks. In analyzing experimental data, researchers often use aggregated results, e.g., average RT for tests or for some types of tasks. However, it was demonstrated that the use of aggregated data leads to biased estimations of associations, effects and differences (e.g., Kievit et al., 2013; Lo & Andrews, 2015; Speelman & McGann, 2013). Mixed effects models allow researchers to analyze an answer on each item within a test for each individual, which may eliminate the majority of the problems associated with the use of aggregated data, e.g., in cases using RT (e.g., Lo & Andrews, 2015). When analyzing tests, MLMs allow researchers to obtain estimations of the sample's mean effect (e.g., NRE) and its values for each participant to disentangle between-individual and within-individual variances in the dependent variable and to estimate the between-level interactions in consideration of the nonindependence of data and different standard errors at each level.

In the current study, we assumed that the data had a hierarchical structure and that the items were processed as nested within individuals. The RT for each item was the dependent variable. To take into account differences in RT between digits and number words, we ran the analysis separately for each format. Before inclusion in the models, all interval variables, including RT, were transformed into Z scores. This step is recommended for multilevel regression analysis with random slope or interaction terms (e.g., Frazier et al., 2004; Hox, Moerbeek, & van de Schoot, 2017). It was shown that if all variables in the model were transformed into Z scores, the obtained regression coefficients could be interpreted as the effect size (e.g., Lorah, 2018).

In the first step, the baseline model without predictors was estimated. The results from this model estimated the predicted sample mean of the RT (in standard deviation) and between-individual and within-individual (between-items) variances in RT. The intraclass correlation coefficient (ICC) was also calculated and indicates the proportion of variance in RT explained by between-individual differences.

Next, in Model 1, the following predictors at the item level were added: 1) the variable "Numerical ratio" ("NR") (was calculated for each item as a smaller number divided by a larger number); 2) the variable "size" (was calculated for each item as the sum of two presented numbers); 3) the variable "accuracy" (0 - an incorrect response and 1 – a correct response; 4) the variable "left position", indicating the position of the larger number (0 if on the right and 1 if on the left); 6) the variable "previous digit", indicating the format of the previous item (0 – for number words comparison and 1 for digits comparison).

The coefficient of variable NR denotes the average NRE for the comparison task. To consider the possible nonlinear association between the numerical ratio and RT, we tested the model with the quadratic term of variable "NR" (Model 2) and compared this model with Model 1. To estimate individual differences in the NRE for digit and number word comparisons, we aimed to test models with random slopes of the variables "NR" and/or "NR squared" (Model 3). This model suggests that the NRE can vary across participants. If the model with a random slope fits better than the model with a fixed slope, it indicates that the NRE significantly varied between individuals. A larger variance corresponds to larger between-individual differences in NRE. In a final step, we added the covariance between the individual slope of NR (and/or NR-squared) and the individual intercept (Model 4), which allowed us to estimate the association between individual average RT and individual NRE. Finally, we calculated the predicted individual value of NRE from Model 3 and Model 4 and estimated the correlation between the NREs for digits and for number words at the individual level.

Results

Descriptive statistics

The average accuracy and RT for the whole test and for the comparison of digits and number words are presented in Table 1.

Table 1

Descriptive statistics for accuracy and RT in digit and number word comparison tasks

Measures	Mean	SD	IQR
Accuracy in digits comparison (% correct answers)	0.90	0.21	0.92; 1.00
Accuracy in number words comparison (% correct answers)	0.89	0.22	0.92; 1.00
RT in digits comparison (msec.)	1323.5	382.2	1125.0; 1526.9
RT in number word comparison (msec.)	1779.2	587.3	1449.6; 2165.5

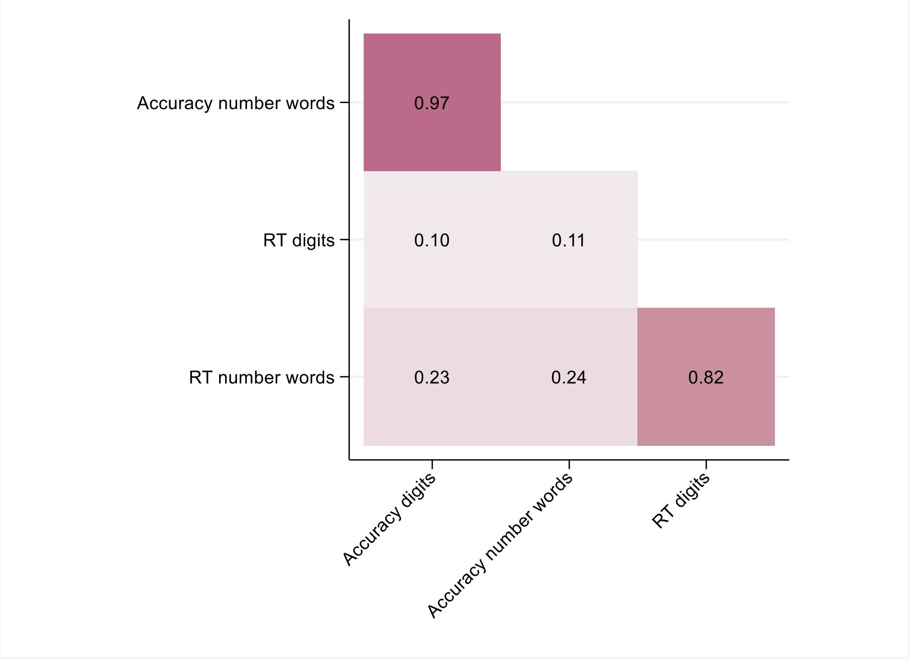
Note: SD – standard deviation; IQR – interquartile range

The descriptive statistics revealed that the accuracies for digit and number word comparisons were high and quite close, while the mean RT was longer for number word comparisons than for digit comparisons.

There are high correlations between accuracies in two symbolic formats and RTs (Fig. 1). At the same time, the correlation between RT and accuracy for each format was low. The correlation between accuracy and RT for digit comparison was positive and weak, while RT and accuracy for number word comparison also had a positive correlation but were larger than those for digit comparison.

Figure 1

Correlations between variables



Results of linear mixed effects models (LMMs)

The results of the LMMs for digits are presented in Table 2.

Table 2

Results of LMMs for RT in the digit comparison test

	Null model	Model 1	Model 2 (non-linear)	Model 3 (random slope)	Model 4 (random slope with covariance)
Fixed effects					
Constant	0.00 (0.02)	-0.08*** (0.02)	-0.05* (0.03)	-0.05* (0.03)	-0.05* (0.03)
NR (Z-scores)		0.07*** (0.005)	0.07*** (0.005)	0.07*** (0.005)	0.07*** (0.005)
NR ^{2}			-0.02*** (0.004)	-0.02*** (0.004)	-0.02*** (0.004)
Size (Z-scores)		-0.04*** (0.004)	-0.03*** (0.004)	-0.03*** (0.004)	-0.03*** (0.004)
Accuracy		0.11*** (0.02)	0.10*** (0.02)	0.10*** (0.02)	0.10*** (0.02)
Left position		-0.12*** (0.009)	-0.12*** (0.009)	-0.12*** (0.009)	-0.12*** (0.009)
Prev. digit		0.10*** (0.01)	0.09*** (0.01)	0.09*** (0.01)	0.09*** (0.01)
Random effects					
Between-individual variance	0.33	0.33	0.33	0.33	0.33
Within-individual variance	0.67	0.66	0.66	0.66	0.66
Variance slope NR				0.000	0.0004
Covariance slope NR and intercept					0.01

Log-likelihood	-41525.71	-41325.37	-41318.70	-41318.70	-41309.81
ICC	0.33	0.33	0.33	0.33	0.33
LR test (Δ df)		400.67*** (5)	13.32*** (1)	0.00 (1)	17.78*** (1)
R-squared (level 1)		0.01	0.01	0.01	0.01
R-squared (level 2)		0.00	0.00	0.00	0.00

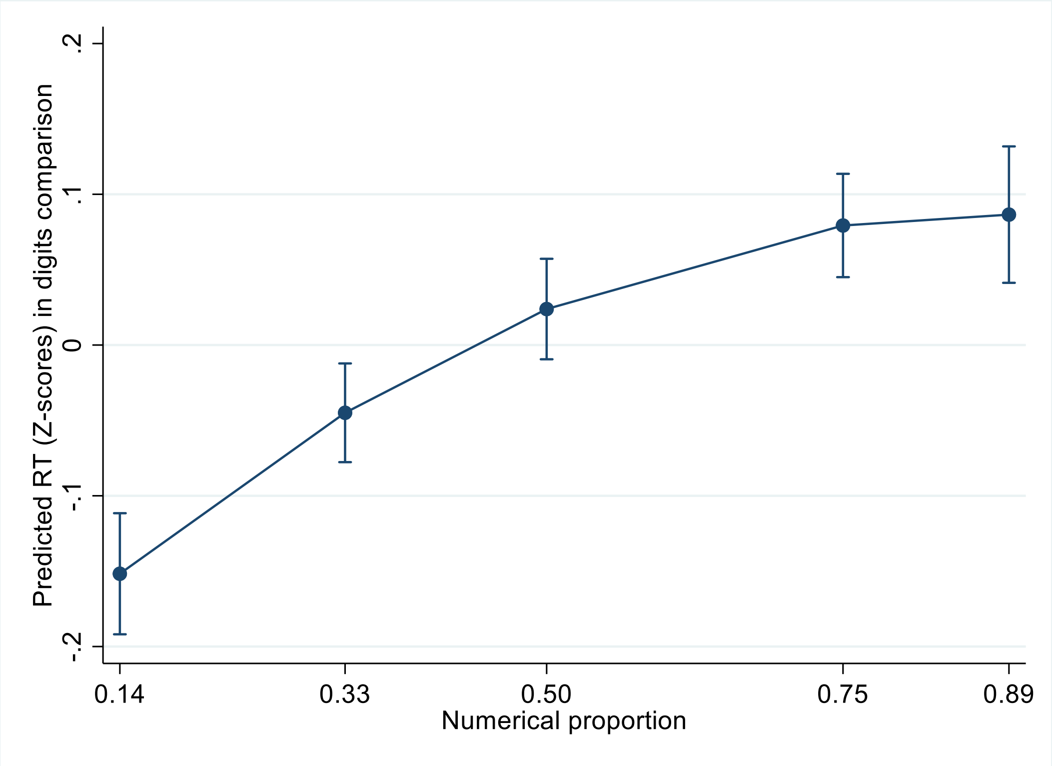
****p*<.001

Note: ICC – intraclass correlation coefficient; LR test – likelihood ratio test; Δ df – differences in degrees of freedom; R-squared at each level was calculated using the formula proposed by Snijders and Boskers (2004)

An analysis revealed that with respect to the digit comparison, the average NRE was significant, but the association between RT and numerical ratio was nonlinear. RT increased when NR increased, but the increase slowed when NR became larger after the numerical proportion was 0.50 (e.g., 4 and 8) (Fig. 2).

Figure 2

Predicted RT (with 95% CI) in digit comparison for different values of numerical proportion

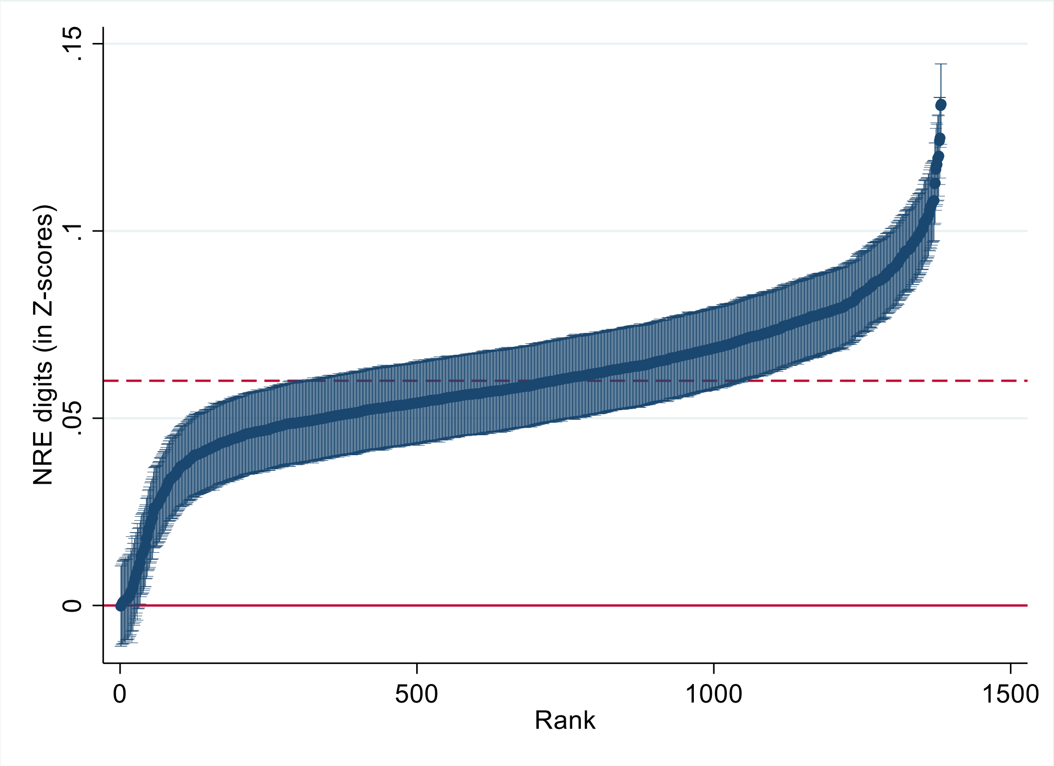


In addition, the association RT with “size” (sum of numbers) was negative, indicating that when the values of numbers increased, the RT decreased (controlling for NR). The RT was larger in correct answers and when the previous item was digit comparison. The RT decreased when the larger number was on the left.

The model with a random slope for the NR variables (Model 3) did not fit the data better than the model with a fixed slope. This means that there were no between-individual differences in the NRE for digit comparison. However, when we added the covariance between random effects for slope and intercept, the model’s fit improved (Model 4). Moreover, in this model, significant interindividual differences in the NRE were identified (Fig. 3).

Figure 3

Individual differences in NRE for digits (from model with random slope and covariance between slope and intercept)



Note: dashed reference line indicates the sample average NRE; solid reference line indicates zero effect

However, this variability was significant because the correlation between random effects for slope and intercept was 1.00. In other words, the variance in the NRE was fully explained by the variance in RT.

It should also be noted that NR, size, accuracy and other variables explained very little of the variability in RTs for the digit comparison task.

Next, we ran mixed effect models for number word comparison with the RTs transformed into Z scores. The results are presented in Table 3.

Table 3

Results of LMMs for RT on number words comparison tasks

Variables	Null model	Model 1	Model 2 (non-linear)	Model 3 (random slope)	Model 4 (random slope with covariance)
Fixed effects					
Constant	0.01 (0.02)	-0.17*** (0.03)	-0.26*** (0.03)	-0.26*** (0.03)	-0.25*** (0.03)
NR (Z-scores)		0.05*** (0.005)	0.05*** (0.005)	0.05*** (0.005)	0.05*** (0.005)
NR ^{2}			0.07*** (0.004)	0.07*** (0.004)	0.07*** (0.004)
Size (Z-scores)		0.08*** (0.004)	0.08*** (0.004)	0.08*** (0.004)	0.08*** (0.004)
Accuracy (1 – correct)		0.07*** (0.02)	0.09*** (0.02)	0.09*** (0.02)	0.09*** (0.02)
Left position		-0.01 (0.008)	-0.002 (0.008)	-0.001 (0.008)	-0.002 (0.008)
Prev. digit		0.20*** (0.008)	0.22*** (0.009)	0.22*** (0.009)	0.22*** (0.009)
		Random effects			
Between-individual variance	0.48	0.48	0.48	0.48	0.44
Within-individual variance	0.52	0.51	0.50	0.50	0.50
Variance slope NR^2				0.002	0.001
Covariance slope NR^2 and intercept					0.02

Log-likelihood	-37770.66	-37264.36	-37143.56	-37141.05	-37118.30
ICC	0.48	0.48	0.49	0.49	0.49
LR test (df)		1012.59*** (5)	241.59*** (1)	5.05* (1)	
R-squared level 1		0.02	0.03	0.03	0.03
R-squared level 2		0.01	0.01	0.01	0.01

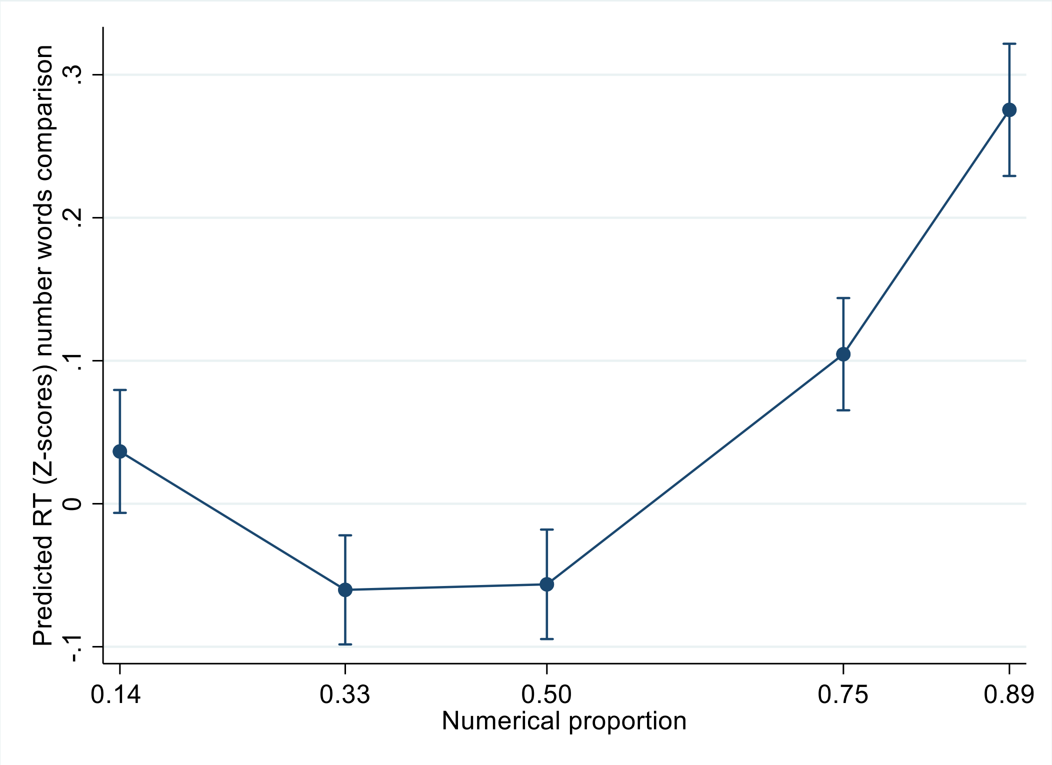
***p<.001

Note: ICC – intraclass correlation coefficient; LR test – likelihood ratio test; Δ df – differences in degrees of freedom; R-squared at each level was calculated using the formula proposed by Snijders and Boskers (2004)

The results of the models for the number words comparison revealed that the NRE was also significant. However, the pattern of changes in RT with increasing numerical proportion was different. At a certain range of numerical proportions (from 0.14 to 0.50), RT might decrease with increasing proportion. However, after this value, RT increased with numerical proportion, which reproduces the classical NRE (Fig. 4).

Figure 4

Predicted RT (with 95% CI) in number words comparison for different values of numerical proportion

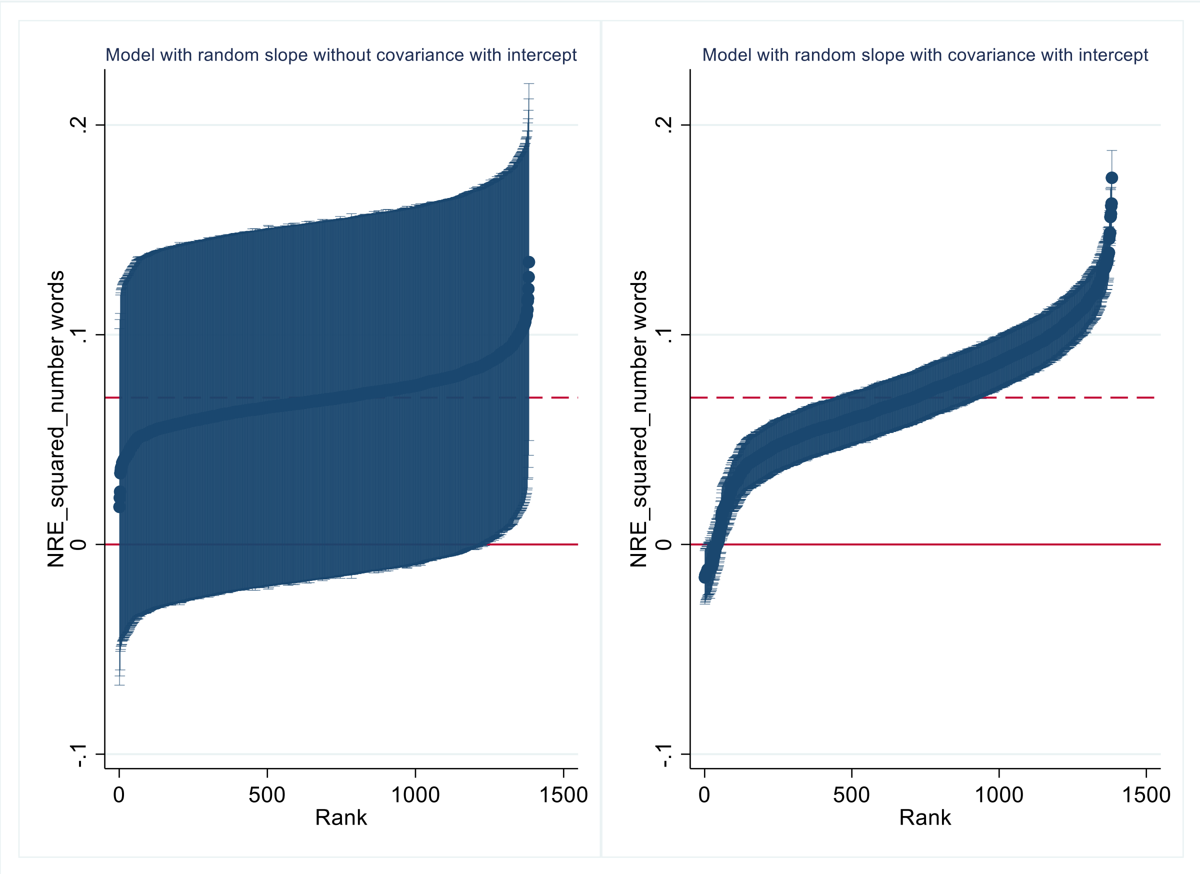


The "size" effect was significant and positive, in contrast with digit comparison, but the position of the larger number on the left was not significant for the number words comparison.

The model with a random slope of the variable NR-squared fitted the data better than the model with a fixed slope. It should also be noted that the slope of the NR-squared variable exhibited significant variance, while the slope for the NR variable did not exhibit significant variance. The distribution of individual slopes of NR-squared from Model 3 is presented in Figure 5 (left side). Although the variance of the slope for the NR-squared variable was significant, the analysis showed that the 95% CI of slopes for all participants included an average sample mean. This result indicates that individual differences in the NRE for number words comparison were tiny.

Figure 5

Between-individual variability in slope for the NR-squared variable in the number words comparison (with 95% CI) in model with random slope (without and with covariance with intercept)



Note: dashed reference line indicates the sample average NRE; solid reference line indicates zero effect

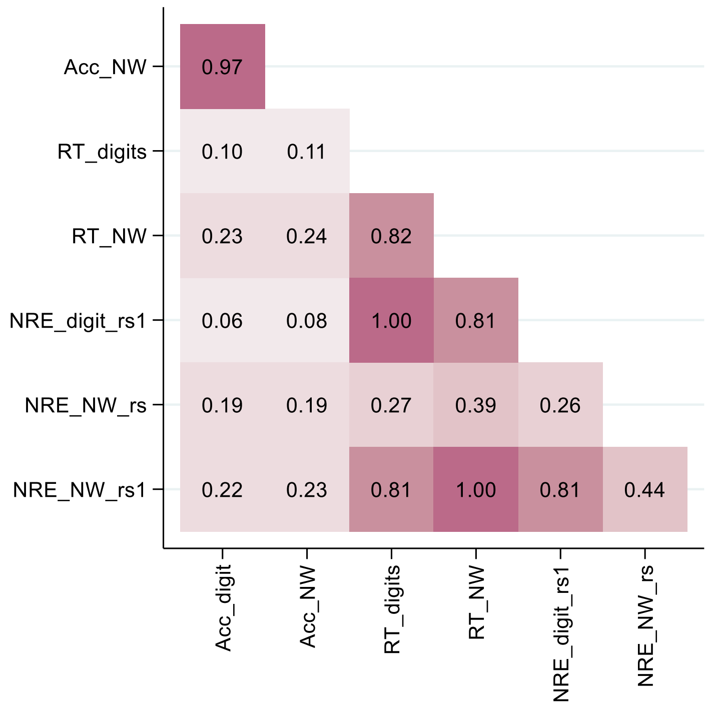
However, when the covariance between the random effects for intercept and slope was added, the between-individual variability of the slope for the NR-squared variable became significant. It should be noted that the standard error for the random effect in the model without covariance (Fig. 5, left side) was greater than that in the model with covariance (Fig. 5, right side).

As in the case of the digit comparison, the variability in the NRE for number words was fully explained by the variability in RT in the model with covariance. The correlation between random effects for slope and intercept was 1.00.

We also tested the correlations between accuracy, mean RT and NRE (from two models) for digits and number words (Fig. 6).

Figure 6

Correlation between measures



Note: Acc_digit – accuracy in digit comparison; Acc_NW – accuracy in number words comparison; RT_digits – mean RT in digits comparison; RT_NW – mean RT in number words comparison; NRE_digit_rs1 – NRE for digits comparison calculated from model with random slope and covariance between slope and intercept; NRE_NW_rs – NRE calculated from model with random slope without covariance; NRE_NW_rs1 - NRE for number words comparison calculated from model with random slope and covariance between slope and intercept.

A correlation analysis demonstrated that the NRE for digits and number words, calculated from the model with random slope and covariance between intercept and slope, was the same as the correlation between the average RT on the two tests. We could not estimate the correlation between the NRE in digits and number words calculated from the model without covariance between intercept and slope because slope variance was insignificant for the digit comparison task and the value of the NRE was the same for all participants.

Discussion

In many studies, the NRE was assumed to be the core feature of numerical representation, reflecting the property of the mental number line and manifested by increasing RT while comparing numbers that are close to each other on the mental number line (e.g., Holloway &

Ansari, 2009). Several studies have also used the individual value of the NRE as an indicator of the precision of numerical representation in symbolic or nonsymbolic formats, with smaller values of NRE corresponding to more precise representation (e.g., Askenazi et al., 2009; De Smedt et al., 2009). However, it has been suggested that NRE can be small or absent due to factors other than precise representation (e.g., Lyons et al., 2015). In addition, it has been doubtful that NRE can be a valid measure of individual precision of numerosity representation because of low reliability (Maloney et al., 2010). Previous studies mostly explored NRE for digits and nonsymbolic formats, while NRE for number words was not well studied.

The current study aimed to investigate individual differences in the NRE for digits and number words. We used mixed-effects models to estimate the sample mean NRE (fixed effect) and the between-individual variability for this effect (random effect for slope) in a large sample of third graders. The estimation of between-individual differences in the NRE is important from the perspective of studies of individual differences, particularly with respect to developmental studies or in studies of associations between NRE and math achievement.

The estimation of fixed effects for the number comparison task in the digit and number word formats revealed that the sample's average NRE was significant for both formats. However, the analysis demonstrated that patterns of changes in RT with increasing numerical ratio were different for digits and number words. For digit comparison, the association between RT and numerical ratio was nonlinear. RT increased when the numerical ratio increased, but growth in RT slowed down gradually. After a ratio of approximately 0.75, the RT did not increase when the numerical ratio increased. This means that the most prominent increase in RT is identified when the numerical ratio changes from small to medium, while the difference in RT between medium and large ratios is tiny. Another pattern was identified for number words comparison. RT did not change when the numerical ratio changed from small to medium, but then, RT became larger (above a numerical ratio of 0.5) and demonstrated fast growth.

These results contradicted some previous findings that concluded the NRE was greater for digits than for number words (e.g., Kadosh et al., 2008). For example, Kadosh and colleagues (2008) determined that the distance effect was smaller for written number words in Hebrew than for digits. However, the replication of that study with Turkish- and English-speaking samples revealed a greater distance effect for number words than for digits, but only for Turkish number words; the distance effect for number words in English showed features similar to those of Hebrew number words (Lukas et al., 2014). It is possible that differences in results are associated with differences in ratios in different studies. As we can see, for small and medium numerical proportions, RT grew faster for digits, while for larger numerical proportions, RT grew faster for number words.

Other obtained results also showed differences in processing digits and number words. Although the estimation of the "size" effect was not the aim of our study, we included the variable "size", reflecting the sum of two compared numbers. The "size" effect reflects the increase in RT in processing large numbers compared to small numbers, holding the proportion (distance) between numbers constant (e.g., Krajcsi, Lengyel, & Kojouharova, 2016). Although the NRE and "size" effect are considered as indicators of internal numerosity representation, some scholars have assumed that the NRE and "size" effect originate from different sources (Verguts & Van Opstal, 2005). Our analysis demonstrated that the "size" effect was negative for digits and positive for number words comparison.

In general, the difference in patterns of the NRE between digits and number words might indicate that differences in processing digits and number words exist. Whether the processing of numerosity varies for digits and number words has been discussed by many scholars. The abstract code model postulates that the processing of numerosity does not depend on the format, digit or number words, as inputs in both symbolic formats should be converted into a common abstract representation of numerosity (e.g., McCloskey, 1992). The triple-code model postulates that the differences between two symbolic formats might depend on the task. Inputs from two symbolic formats should be converted into one common abstract representation but only for operations that require the understanding of the meaning of numerosity, for example, for numerosity comparison or matching tasks (Dehaene & Cohen, 1997; Dehaene & Akhavein, 1995). For other operations, e.g., for calculations, the digit (visual) inputs should be transformed into number word (verbal) format. Encoding complex and multiple representation models assume that each symbolic format has a separate abstract representation; thus, the processing of digits and number words is different in any task (Campbell & Clark, 1992; Campbell & Epp, 2004; Cohen, Warren, & Blanc-Goldhammer, 2013; Fias, 2001). As we used a comparison task and obtained different patterns of NREs for digits and number words, we might conclude that the obtained results are better fitted to the Multiple Representation hypothesis (e.g., Cohen et al., 2013).

Regarding individual differences in the NRE, our analysis revealed that a significant between-individual variance in the NRE was fully explained by the variance in individual RT. In other words, the longer the RT is, the greater the NRE. If we assumed that the individual value of the NRE did not depend on the individual RT, the between-individual variance in the NRE would be insignificant, and the model with no covariance between the NRE and the individual RT fit the data worse than the model with such covariance. Hence, we should assume that the individual NRE is strongly correlated with the individual RT and that this correlation provides between-individual variability in the NRE. However, if the value of the NRE reflects the speed of numerical comparison, there is no reason to use the NRE, as an average or median RT can be used instead. In addition, some studies have demonstrated that the reliability of the RT is greater than the reliability of the difference in the RT between the two conditions (e.g., Caruso, 2004; Edwards, 2001).

These results are in line with the current discussion regarding the inappropriate use of experimental tasks in studies of individual differences (e.g., Borsboom et al., 2009; Hedge et al., 2018; Rouder & Haaf, 2019). Our study revealed that the NRE is robust and may reflect an important feature of numerical representation. However, NRE does not significantly vary across individuals, at least in comparison tasks with one-digit numbers among third graders. Consequently, the NRE produced in a one-digit comparison task does not reflect the level of individual ability and cannot be used in the investigation of individual differences in numerosity representation without testing the significance of between-individual differences.

Reference

Ashkenazi, S., Mark-Zigdon, N., & Henik, A. (2009). Numerical distance effect in developmental dyscalculia. *Atypical Development of Numerical Cognition*, *24*(4), 387–400. https://doi.org/10.1016/j.cogdev.2009.09.006

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12-28.

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

Bernardo, A. B. I. (2001). Asymmetric activation of number codes in bilinguals: Further evidence for the encoding complex model of number processing. *Memory & Cognition*, *29*(7), 968–976. https://doi.org/10.3758/BF03195759

Borsboom, D., Kievit, R. A., Cervone, D., & Hood, S. B. (2009). The Two Disciplines of Scientific Psychology, or: The Disunity of Psychology as a Working Hypothesis. In J. Valsiner, P. C. M. Molenaar, M. C. D. P. Lyra, & N. Chaudhary (eds.), *Dynamic Process Methodology in the Social and Developmental Sciences* (pp. 67–97). Springer US. https://doi.org/10.1007/978-0-387-95922-1_4

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*(3), 389–411. https://doi.org/10.1037/met0000159

Campbell, J. I. D., & Clark, J. M. (1992). Cognitive Number Processing: An Encoding-Complex Perspective. В J. I. D. Campbell (Ред.), *Advances in Psychology* (Т. 91, сс. 457–491). North-Holland. https://doi.org/10.1016/S0166-4115(08)60894-8

Campbell, J. I., & Epp, L. J. (2004). An encoding-complex approach to numerical cognition in Chinese-English bilinguals. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *58*(4), 229.

Caruso, J. C. (2004). A Comparison of the Reliabilities of Four Types of Difference Scores for Five Cognitive Assessment Batteries. *European Journal of Psychological Assessment*, *20*(3), 166–171. https://doi.org/10.1027/1015-5759.20.3.166

Cohen, D. J., Warren, E., & Blanc-Goldhammer, D. (2013). Cross-format physical similarity effects and their implications for the numerical cognition architecture. *Cognitive Psychology*, *66*(4), 355–379. https://doi.org/10.1016/j.cogpsych.2013.03.001

Damian, M. F. (2004). Asymmetries in the processing of Arabic digits and number words. *Memory & Cognition*, *32*(1), 164–171. https://doi.org/10.3758/BF03195829

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. https://doi.org/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. *Special Issue: Typical Development of Numerical Cognition*, *103*(4), 469–479. https://doi.org/10.1016/j.jecp.2009.01.010

Dehaene, S., & Akhavein, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 314–326. https://doi.org/10.1037/0278-7393.21.2.314

Dehaene, S., & Cohen, L. (1997). Cerebral Pathways for Calculation: Double Dissociation between Rote Verbal and Quantitative Knowledge of Arithmetic. *Cortex*, *33*(2), 219–250. https://doi.org/10.1016/S0010-9452(08)70002-9

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

Dietrich, J. F., Huber, S., Klein, E., Willmes, K., Pixner, S., & Moeller, K. (2016). A Systematic Investigation of Accuracy and Response Time Based Measures Used to Index ANS Acuity. *PLOS ONE*, *11*(9), e0163076. https://doi.org/10.1371/journal.pone.0163076

Edwards, J. R. (2001). Ten Difference Score Myths. *Organizational Research Methods*, *4*(3), 265–287. https://doi.org/10.1177/109442810143005

Fias, W. (2001). Two routes for the processing of verbal numbers: Evidence from the SNARC effect. *Psychological Research*, *65*(4), 250–259. https://doi.org/10.1007/s004260100065

Field, A. P., & Wright, D. B. (2011). A Primer on Using Multilevel Models in Clinical and Experimental Psychopathology Research. *Journal of Experimental Psychopathology*, *2*(2), 271–293. https://doi.org/10.5127/jep.013711

Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing Moderator and Mediator Effects in Counseling Psychology Research. *Journal of Counseling Psychology*, *51*(1), 115–134. https://doi.org/10.1037/0022-0167.51.1.115

Göbel, S. M., Watson, S. E., Lervåg, A., & Hulme, C. (2014). Children's Arithmetic Development: It Is Number Knowledge, Not the Approximate Number Sense, That Counts. *Psychological Science*, *25*(3), 789–798. https://doi.org/10.1177/0956797613516471

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

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and evaluation in counseling and development*, *34*(3), 177–189.

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, *39*(1), 101–117. https://doi.org/10.3758/BF03192848

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*(1), 17–29. https://doi.org/10.1016/j.jecp.2008.04.001

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2017). *Multilevel Analysis: Techniques and Applications* (3-е изд.). Routledge. https://doi.org/10.4324/9781315650982

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

Ivanova, A., Kardanova, E., Merrell, C., Tymms, P., & Hawker, D. (2018). Checking the possibility of equating a mathematics assessment between Russia, Scotland and England for children starting school. *Assessment in Education: Principles, Policy & Practice*, *25*(2), 141–159. https://doi.org/10.1080/0969594X.2016.1231110

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*(3), 1221–1247. https://doi.org/10.1016/j.cognition.2007.06.004

Kadosh, R. C., Henik, A., & Rubinsten, O. (2008). Are Arabic and verbal numbers processed in different ways? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1377–1391. https://doi.org/10.1037/a0013413

Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00513

Kolkman, M. E., Kroesbergen, E. H., & Leseman, P. P. M. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learning and Instruction*, *25*, 95–103. https://doi.org/10.1016/j.learninstruc.2012.12.001

Krajcsi, A., Lengyel, G., & Kojouharova, P. (2016). The Source of the Symbolic Numerical Distance and Size Effects. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01795

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01171

Lonnemann, J., Linkersdörfer, J., Hasselhorn, M., & Lindberg, S. (2011). Symbolic and non-symbolic distance effects in children and their connection with arithmetic skills. *Neurocognitive Precursors of Difficulties in Reading and Arithmetic*, *24*(5), 583–591. https://doi.org/10.1016/j.jneuroling.2011.02.004

Lorah, J. (2018). Effect size measures for multilevel models: Definition, interpretation, and TIMSS example. *Large-Scale Assessments in Education*, *6*(1), 8. https://doi.org/10.1186/s40536-018-0061-2

Lukas, S., Krinzinger, H., Koch, I., & Willmes, K. (2014). Number Representation: A Question of Look? The Distance Effect in Comparison of English and Turkish Number Words. *Quarterly Journal of Experimental Psychology*, *67*(2), 260–270. https://doi.org/10.1080/17470218.2013.802002

Lyons, I. M., Nuerk, H.-C., & Ansari, D. (2015). Rethinking the implications of numerical ratio effects for understanding the development of representational precision and numerical processing across formats. *Journal of Experimental Psychology: General*, *144*(5), 1021–1035. https://doi.org/10.1037/xge0000094

Maloney, E. A., Risko, E. F., Preston, F., Ansari, D., & Fugelsang, J. (2010). Challenging the reliability and validity of cognitive measures: The case of the numerical distance effect. *Acta Psychologica*, *134*(2), 154–161. https://doi.org/10.1016/j.actpsy.2010.01.006

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*, *44*(1–2), 107–157. https://doi.org/10.1016/0010-0277(92)90052-J

Mundy, E., & Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *Special Issue: Typical Development of Numerical Cognition*, *103*(4), 490–502. https://doi.org/10.1016/j.jecp.2009.02.003

Nieder, A., & Miller, E. K. (2003). Coding of Cognitive Magnitude. *Neuron*, *37*(1), 149–157. https://doi.org/10.1016/S0896-6273(02)01144-3

Nuerk, H. C., Moeller, K., Klein, E., Willmes, K., & Fischer, M. H. (2011). Extending the mental number line: A review of multi-digit number processing. *Zeitschrift für Psychologie/Journal of Psychology*, *219*(1), 3.

Nuerk, H.-C., Kaufmann, L., Zoppoth, S., & Willmes, K. (2004). On the Development of the Mental Number Line: More, Less, or Never Holistic With Increasing Age? *Developmental Psychology*, *40*(6), 1199–1211. https://doi.org/10.1037/0012-1649.40.6.1199

Núñez-Peña, M. I., & Suárez-Pellicioni, M. (2014). Less precise representation of numerical magnitude in high math-anxious individuals: An ERP study of the size and distance effects. *Biological Psychology*, *103*, 176–183. https://doi.org/10.1016/j.biopsycho.2014.09.004

Ponterotto, J. G., & Ruckdeschel, D. E. (2007). An overview of coefficient alpha and a reliability matrix for estimating adequacy of internal consistency coefficients with psychological research measures. *Perceptual and motor skills*, *105*(3), 997–1014.

Price, G. R., Palmer, D., Battista, C., & Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: Reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychologica*, *140*(1), 50–57. https://doi.org/10.1016/j.actpsy.2012.02.008

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

Rousselle, L., & Noël, M. P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. *Cognition*, *102*(3), 361-395. https://doi.org/10.1016/j.cognition.2006.01.005

Sasanguie, D., & Reynvoet, B. (2014). Adults' Arithmetic Builds on Fast and Automatic Processing of Arabic Digits: Evidence from an Audiovisual Matching Paradigm. *PLOS ONE*, *9*(2), e87739. https://doi.org/10.1371/journal.pone.0087739

Sasanguie, D., Defever, E., Van den Bussche, E., & Reynvoet, B. (2011). The reliability of and the relation between non-symbolic numerical distance effects in comparison, same-different judgments and priming. *Acta Psychologica*, *136*(1), 73–80. https://doi.org/10.1016/j.actpsy.2010.10.004

Speelman, C. P., & McGann, M. (2013). How Mean is the Mean? *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00451

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. https://doi.org/10.3758/PBR.15.2.419

Verguts, T., & Van Opstal, F. (2005). Dissociation of the distance effect and size effect in one-digit numbers. *Psychonomic Bulletin & Review*, *12*(5), 925–930. https://doi.org/10.3758/BF03196787

Zhang, J., & Norman, D. A. (1995). A representational analysis of numeration systems. *Cognition*, *57*(3), 271–295. https://doi.org/10.1016/0010-0277(95)00674-3

Yulia Kuzmina

Center for Psychometrics and Measurement in Education, Institute of Education, National Research University Higher School of Economics, research fellow, e-mail (jkuzmina@hse.ru)