



**QUEEN'S
UNIVERSITY
BELFAST**

The stability of individual differences in basic mathematics-related skills in young children at the start of formal education

O'Connor, P., Morsanyi, K., & McCormack, T. (2019). The stability of individual differences in basic mathematics-related skills in young children at the start of formal education. *Mind, Brain, and Education*. <https://doi.org/10.1111/mbe.12190>

Published in:
Mind, Brain, and Education

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights
Copyright 2019 Wiley. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Open Access
This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: <http://go.qub.ac.uk/oa-feedback>

Stability of basic mathematics-related skills

Running head: Stability of basic mathematics-related skills

**The stability of individual differences in basic mathematics-related skills in young
children at the start of formal education**

Word count: 5441 excluding references and tables

The authors declare that there are no potential sources of conflict of interest

Abstract

The current study investigated the development of children's performance on tasks that have been suggested to underlie early mathematics skills, including measures of cardinality, ordinality and intelligence. Eighty-seven children were tested in their first (T1) and second (T2) school year (at age 5 and 6). Children's performance on all tasks demonstrated good reliability and significantly improved with age. Correlational analyses revealed that performance on some mathematics-related tasks were non-significantly correlated between T1 and T2 (Number line and Number comparison), showing that these skills are relatively unstable. Detailed analyses also indicated that the way children solve these tasks show qualitative changes over time. By contrast, children's performance on measures of intelligence and non-numerical ordering abilities were strongly correlated between T1 and T2. Additionally, ordering skills also showed moderate to strong correlations with counting procedures both cross-sectionally and longitudinally. These results suggest that, initially, mathematics skills strongly rely on non-mathematical abilities.

Key Words: Longitudinal Design; Magnitude Processing; Mathematics Development; Stability; Mental Number Line; Order Processing

Introduction

Much research into important early predictors of mathematical development have been concerned with investigating the extent to which cardinal skills (both symbolic and non-symbolic) are involved in the acquisition of symbolic number knowledge. However, recent evidence has suggested that ordinality, a key property of numbers along with cardinality (Gelman & Gallistel, 1978), may also play an important role in early number development, as may other general factors, such as intelligence (e.g., Roth et al., 2015). Studies often administer a set of these predictors together, in order to establish which measures are most closely related to formal mathematics skills. Such results are then interpreted in a causal way, by assuming that tasks that most strongly predict mathematics skills form the foundation of those skills. Nevertheless, these studies typically do not investigate and take into account the stability of individual differences in those predictors over time. Indeed, it is important to distinguish between tasks that are merely correlated with mathematics skills, but do not reflect stable individual differences, as opposed to skills that show intra-individual stability, and might form the foundations of learning maths, as this knowledge could inform theories regarding the typical and atypical development of maths skills, as well as intervention efforts. Thus, the aim of the present study was to investigate the stability of several numerical and non-numerical predictors of early mathematical success, in order to identify a set of skills that are already in place at the start of formal education, and could support the development of early mathematics competence. In the following sections, we describe the tasks that we considered in our study.

Previous research has proposed the existence of an innate, evolutionarily ancient system for processing cardinality, which is not dependent on language skills and is not unique to humans, referred to as the ‘Approximate Number System’ (ANS; Piazza et al., 2010). Evidence from meta-analyses appear to support the role of the ANS in early numerical

development (Chen & Li, 2014; Fazio, Thompson, Bailey & Siegler, 2014; Schneider et al., 2017). Many studies have used the Dot comparison task as a measure of the ANS; however, there have been questions raised about the reliability and validity of this measure (e.g. Gilmore et al., 2013; Inglis & Gilmore, 2013, 2014; Maloney, Risko, Preston, Ansari & Fugelsang, 2010; Price, Palmer, Battista & Ansari, 2012). There is evidence to suggest that a Non-symbolic addition task (e.g., Barth, La Mont, Lipton & Spelke, 2005) may be a more viable alternative measure of the ANS. Several studies (Barth, Beckmann & Spelke, 2008; Barth et al., 2005; Li et al., 2017) have found that young children can perform above chance on this task and show similar patterns of responding as adults. Additionally, this task appears to measure a similar underlying construct as the Dot comparison task (Gilmore et al., 2014). Performance on the Non-symbolic addition task has also been linked to mathematical achievement in developmental studies (Gilmore, McCarthy & Spelke, 2010; Wong, Ho & Tang, 2010).

Whereas much research has focussed on non-symbolic magnitude skills, symbolic magnitude skills (typically indexed by the Number comparison task) appear to be more strongly related to maths. This is particularly evident in studies involving children aged six and over (De Smedt, Noël, Gilmore & Ansari, 2013; Fazio et al., 2014; Schneider et al., 2017). This suggests that these skills may exert more of an influence after children have had some experience of formal maths learning at school.

There is now considerable evidence in support of a link between order-processing skills and mathematical abilities. Although it appears that numerical ordering skills become particularly important from around the age of six or seven (Attout & Majerus, 2018; Lyons et al., 2014; Sasanguie & Vos, 2018), there is also now emerging evidence in support of a role of non-numerical ordering in mathematical development in the case of younger children (Attout, Noël & Majerus, 2014; Morsanyi, van Bers, O'Connor & McCormack, 2018;

O'Connor, Morsanyi & McCormack, 2018). Non-numerical order processing measures can be broadly divided into two categories: those involving the retrieval of a familiar sequence from long-term memory, such as the order of familiar daily events, familiar everyday sequences, the months of the year or letters (Morsanyi, O'Mahony, & McCormack, 2017; O'Connor et al., 2018; Sasanguie et al., 2017; Vos, Sasanguie, Gevers & Reynvoet, 2017), and those involving the retrieval of a novel, arbitrary sequence from short-term memory (Order WM task; Attout & Majerus, 2015; 2018; Attout et al., 2014). O'Connor et al. (2018) found that both numerical and non-numerical ordering measures were related to early mathematical achievement in 4-5-year-old children. However, the ordering of familiar sequences was the strongest predictor of children's maths achievement at the end of their first year of school, and also longitudinally predicted maths achievement at the end of their second year, after controlling for the effect of several numerical and non-numerical tasks. Order WM performance also correlated with maths achievement but did not explain additional variance after taking into account the effect of the order-processing measures involving the retrieval of familiar content.

Performance on Number line estimation tasks have been proposed to reflect representations of number along a mental number line (e.g., Bonato, Zorzi, & Umiltà, 2012; Kaufman, Vogel, Starke, Kremser, & Schocke, 2009; Link, Huber, Nuerk & Moeller, 2014; Moyer & Landauer, 1967). Furthermore, Number line performance has been linked to early numerical development, even in very young children (e.g. Berteletti, Lucangeli, Piazza, Dehaene & Zorzi, 2010; Siegler & Booth, 2004; Siegler & Opfer, 2003), suggesting that estimation skills may be important to early maths learning. However, Schneider et al. (2018) found that effect sizes, regarding the relation between Number line estimation and maths achievement, increased with age (.30 for children below 6; .44 for 6-9 year olds and .49 for 9-

14-year-olds). This finding suggests a co-development of Number line estimation and formal mathematical skills with time.

Finally, there is evidence that intelligence is strongly related to academic achievement (e.g., Morsanyi, van Bers, McCormack & McGourty, 2018) , and also to other socio-economically relevant outcomes, such as future employment and income (Deary, Strand, Smith & Fernandes, 2007; Roth et al., 2015; Strenze, 2007). This suggests not only that verbal and non-verbal intelligence are important to mathematical achievement during the school years, but are also related to children's future prospects after they have left school.

As mentioned above, whilst many studies have tried to identify which early predictors are important to numerical development, there is a relative lack of understanding as to the stability of these skills over time. Nevertheless, there are a few longitudinal studies that reported data regarding the stability of performance on some basic numerical tasks (e.g. Attout et al., 2014; Reeve, Reynolds, Humberstone & Butterworth, 2012; Xenidou-Dervou, Molenaar, Ansari, van der Schoot & van Lieshout, 2017), the findings of which are summarized in Table 1. These results suggest that although performance on these tasks shows some consistency over time during the first school years, the strength of the correlations is generally weak to moderate, and sometimes even non-significant.

[INSERT TABLE 1 HERE]

The current study

The aim of the current study was to investigate the development of ordinal, cardinal and intelligence measures between children's first (T1) and second (T2) year of primary school (i.e., between the ages of four and six) in a group of pupils from Northern Ireland, a country which has one of the youngest school starting ages in Europe (Eurydice at NFER,

2012), as children in Northern Ireland begin primary school on the first September after their fourth birthday. We expected that each of these skills would show significant improvements between the first two years of school. Additionally, we aimed to assess the stability of individual differences in these skills, by analyzing correlations between performance on each task across the two school years. Strong correlations between task performance at each time point would show evidence of the stability of the skills that the tasks measure, whilst weak or non-significant correlations would suggest that the underlying processes and strategies are still developing.

We also investigated the relations between each task and counting skills (i.e., counting forward and backward from different starting points), both cross-sectionally and longitudinally. We considered the counting task an indicator of familiarity with the number system, as well as the flexibility with which children could use numbers. These skills are essential for the development of arithmetic abilities (e.g., Geary, Brown, & Samaranayake, 1991; Lemaire & Siegler, 1995), and, thus, are very important for the development of formal mathematics skills. Tasks that show strong relations with counting skills at each time point can be considered diagnostic of mathematics skills. Nevertheless, we expected that the way children perform some of these tasks might change considerably during the first school years (for example, because children automatize some procedures or develop new strategies). Only tasks that show stability over time can be considered as potential candidates for skills that might form the foundations of mathematics abilities. Indeed, if performance on a task is not strongly related to performance on the same task at a later time point, it could not be considered as a good indicator of individual differences in a basic skill.

Where it was possible, we also performed detailed analyses of some within-task variables, in order to investigate how the effect of these variables on children's performance

changed between T1 and T2. This could help in explaining the findings regarding the stability of individual differences on each task.

Method

Participants¹

Eighty-seven children participated in the study (43 females, Mean age at T1 = 4 years 11 months; $SD = 3.73$ months. Mean age at T2 = 6 years 2 months, $SD = 3.44$ months). Due to the demographics of the population in Northern Ireland, the vast majority of children were of Caucasian origin. Children's level of socio-economic deprivation was determined using the Northern Ireland Multiple Deprivation Measure (Northern Ireland Statistics and Research Agency, 2010). This measure assigns a deprivation score to each electoral ward in Northern Ireland based on seven indices (income deprivation; employment deprivation; health deprivation and disability; education, skills and training deprivation; proximity to services; living environment; crime and disorder). A higher score indicates a higher level of deprivation for the area. The scores can be interpreted as percentiles (e.g., a score of 10 means that the area is less deprived than 90% of all postcode-based areas within Northern Ireland). In the current sample, deprivation scores ranged from 1.85 – 68.57 (Median deprivation score = 11.00). Based on children's postcodes, most children came from areas with very low levels of socio-economic deprivation, reflecting that the majority of the sample came from areas of higher socio-economic status, although deprivation indices ranged from low to medium. One child did not provide a postcode, so a deprivation score could not be calculated for them.

¹ The children in the current study were the same as in O'Connor et al. (2018) and some results regarding those children's performance at T1 are also reported in that paper.

Materials

IQ. Verbal and non-verbal intelligence was measured using the Vocabulary and Block Design subtests of the Wechsler Preschool & Primary Scale of Intelligence - Third UK Edition (WPPSI-III UK; Wechsler, 2003). Children's estimated full-scale IQ scores were computed following the method outlined in Sattler and Dumont (2004) and were found to be within the normal range at both time points. (Mean T1 IQ score = 95.92, $SD = 13.51$; Mean T2 IQ score = 101.80, $SD = 12.45$).

Order Working Memory (WM) task. This was based on a similar measure developed by Majerus and colleagues (Majerus, Poncelet, Greffe & Van der Linden, 2006). This task measures children's ability to retain and manipulate serial order information by measuring their ability to recreate the correct sequence of a list of animal names that were presented to them through a set of earphones, using cards depicting the animals. The length of item sequences ranged from 2 to 7 with 4 items at each level. Split-half reliability estimates, using the Spearman-Brown formula, indicated good reliability (T1: $r = .93$; T2: $r = .95$).

Daily Events task. This computerised task was based on Friedman's (1977, 1990) temporal ordering task. Children were shown three daily events, out of a set of six (half of the trials were in canonical order, from left to right; half were in a mixed order) and judged whether the order was correct or not, from right to left, by pressing either a tick or a cross on the touchscreen monitor. Since each trial was presented twice (with a total of 24 trials), a split-half reliability was calculated using the Spearman-Brown coefficient, which was found to be adequate (T1: $r = .57$. T2: $r = .76$).

Counting. In this task, children were first asked to count to 50 (T1) or 100 (T2) twice. Additionally, children had to count forwards and backwards (three trials for each direction) from different starting points. Children could correct themselves once during any trial and were stopped once they had correctly recited the next three numbers in the sequence. A score

Stability of basic mathematics-related skills

of 1 was given for each trial in which children correctly recited the next three numbers in the sequence. A total counting score was calculated by adding z scores for all 3 counting measures (counting until 50 or 100, counting forwards from different starting positions and counting backwards from different starting positions). The reliability estimate for the task was good (T1 Cronbach's Alpha = .77; T2 Cronbach's Alpha = .75).

Non-symbolic Addition. This computerised task was based on the one used by Gilmore, McCarthy and Spelke (2010), in which children had to approximately add two arrays of dots together (sum array) and compare the sum of these to a comparison array of dots. The numerical ratio of the sum and comparison arrays was manipulated across the 24 trials (1:2, 3:5, and 2:3), with 8 trials per ratio. The number of dots for both arrays varied from 6 to 45, with 6 being the lowest number of dots as this reduced the possibility that children could subitize the number of dots presented. Perceptual variables (dot size, density and array size) were also varied, so that they correlated with numerosity on half the trials (congruent trials) and were uncorrelated on the other half of the trials (incongruent trials), reducing the possibility that children may have used perceptual information as a cue when judging which array was the most numerous. In the task, children had to press one of two buttons on the touchscreen to indicate which character they thought had the most marbles. They first completed four practice trials, with feedback given on their performance. Children were given a score of 1 if they correctly judged which character had the most marbles. Reliability for this task was low, but acceptable (T1 Cronbach's alpha = .50; T2 Cronbach's alpha = .63), and one-sample t tests confirmed that children performed above chance at each ratio.

Number comparison. In a computerised task, children were presented with a target number (between 1-4 or 6-9) and were asked to press one of two buttons (either a large square or a small square) to indicate whether they thought that the number on the screen was

Stability of basic mathematics-related skills

larger or smaller than 5. Children were presented with 4 practice trials before completing the task. Task performance showed high reliability (T1 Cronbach's Alpha = .88; T2 Cronbach's Alpha = .84).

Number line task. This was a computerised task in which children had to indicate the position of numbers along a 1-10 and a 1-20 number line, both of which were of equal length (1000 pixels). There were two practice trials and six experimental trials for each scale. Children's error for each individual trial was calculated as the distance in pixels between children's estimated position and the actual position of the target number. The average of children's errors across both 1-10 and 1-20 scales was used as the overall measure of estimation error for the task. The task showed adequate reliability (T1 Cronbach's Alpha = .70; T2 Cronbach's Alpha = .71).

Procedure

The study received ethical approval from the university department's ethics committee. Parents gave consent for their child to take part in the study. In Session 1, all children completed the Number comparison task, the Order WM task and finally, the Non-Symbolic Addition task. In Session 2, children completed the Daily Events Task, followed by the WPPSI-III subtests, Counting task and then finally, the Number line task. In year 1, there was a 3-month gap between session 1 and session 2, whilst in Year 2, there was a 2-month gap between session 1 and session 2. The computer-based tasks were designed using E-Prime Version 2.0. These tasks were presented on a touch screen, connected to a laptop. All tasks were administered individually. For all computer-based tasks, accuracy and reaction times were recorded, but in the following we only report the results regarding accuracy.

Results

The descriptive statistics for children's performance on tasks at T1 and T2, as well as the results of repeated measures t tests, and the effect sizes of differences across the two time points are presented in Table 2. The t tests indicated that children's task accuracy improved significantly on all tasks, when compared to their performance on the same tasks at T1. The effect size of these developmental changes was large for all tasks, with the exception of the Order WM task, where the effect size was medium.

[INSERT TABLE 2 HERE]

Children's performance on the counting task also showed large improvements over this period. At T1, the median number that children were able to count up to (out of a maximum of 50) was 39, whilst at T2, the median number that children were able to count up to (out of 100) was 100, showing that by the second year of school, most children were familiar with the number system up to 100. Additionally, children made significantly fewer mistakes at T2 when they were counting forward and backward from different starting points (T1 accuracy: 76%; T2 accuracy: 92%), although they were given larger starting numbers at T2.

In order to investigate the stability of individual differences on these tasks, correlation analyses were conducted between performance on each task at T1 and T2. A bootstrap procedure (using 10,000 samples) was also applied to obtain 95% confidence intervals for the correlation coefficients. Figure 1 shows that performance on the majority of measures at T1 showed significant bootstrap correlations with performance at T2. The only exceptions were Number comparison; [$r = .20$, 95% CI $(-.01, .34)$], and Number line performance [$r = -.01$, 95% CI $(-.25, .25)$]. The 95% confidence intervals for the Order WM; [$r = .62$, 95% CI $(.47, .74)$], Vocabulary [$r = .61$, 95% CI $(.46, .72)$], and Daily events tasks [$r = .48$, 95% CI $(.30,$

Stability of basic mathematics-related skills

.61)], did not overlap with the 95% confidence intervals for the Number line task, which suggests that the magnitude of correlations for the former three tasks was significantly greater than the correlation for Number line performance. The 95% confidence intervals for the Order WM and Vocabulary measures also did not overlap with the confidence intervals for Number comparison, again indicating that the correlations between T1 and T2 task performance for Order WM and Vocabulary were significantly greater than the correlation between the two time-points for the Number comparison task. That is, there were significant variations in the stability of individual differences on these tasks, with performance on the Number line and Number comparison tasks appearing to be the least stable, and Order WM and Vocabulary performance being the most stable. The 95% confidence intervals for Block design; [$r = .41$, 95% CI (.24, .56)], Counting [$r = .39$, 95% CI (.10, .60)], and Non-symbolic addition tasks [$r = .29$, 95% CI (.05, .48)], overlapped with the 95% confidence intervals for the Number line and Number comparison tasks, which suggests that the magnitude of correlations for the former three tasks was not significantly greater than the correlation for Number line and Number comparison performance. Overall, these results suggested that the way children performed most numerical tasks showed important changes as a result of formal education, whereas individual differences on non-numerical tasks showed moderate to high stability.

[INSERT FIGURE 1 HERE]

A correlation analysis was conducted to investigate the relationships between performance on each task and Counting ability at both time points. As shown in Table 3, Order WM, Daily Events and Number comparison performance showed the most consistent correlations with Counting ability, which was present for both cross-sectional and

longitudinal analyses. Regarding verbal and non-verbal intelligence, Vocabulary skills related to Counting ability at both time points, but not longitudinally, whereas nonverbal intelligence appeared to play a more important role at T2. For Non-symbolic addition, the pattern was similar to non-verbal intelligence, as the task was significantly related to counting only at T2. Number line performance was unrelated to counting ability at both time points.

[INSERT TABLE 3 HERE]

In order to better investigate the developmental changes that happened over the first school years, we conducted further analyses where we investigated the effects of various within-task variables (e.g., distance effects, ratio effects).

Daily events

A 2 x 2 mixed ANOVA was carried out to determine the effect of time (T1 and T2) and trial order (canonical and mixed) on Daily events task accuracy (Figure 2). The analysis revealed a significant main effect of trial order; $F(1, 86) = 37.48, p < .001, \eta_p^2 = .30$. Children performed significantly better on mixed-order trials (Mean = 79%), compared to canonical order trials (Mean = 62%). There was also a main effect of time; $F(1, 86) = 65.20, p < .001, \eta_p^2 = .43$. Children performed significantly better on the task at T2 (Mean = 76%) than at T1 (Mean = 65%). However, there was no interaction between time and trial order ($p = .233$).

[INSERT FIGURE 2 HERE]

Number comparison

A 2 x 4 mixed ANOVA was carried out to determine the effect of time (T1 and T2) and numerical distance between the two to-be-compared numbers (distances of 1, 2, 3 or 4) on Number comparison accuracy. The analysis revealed a significant main effect of time;

Stability of basic mathematics-related skills

$F(1, 87) = 133.12, p < .001, \eta_p^2 = .61$. Children performed significantly better on the task at T2 (Mean = 95%), compared to T1 (Mean = 72%). There was also a main effect of numerical distance; $F(2.78, 241.51) = 25.93, p < .001, \eta_p^2 = .23$. Bonferroni-corrected pairwise comparisons revealed that children performed significantly worse on numerical distance 1 trials (Mean = 77%), compared to performance on distance 2 (Mean = 84%; $p < .001$); distance 3 (Mean = 85%; $p < .001$); and distance 4 trials (Mean = 88%; $p < .001$). The analysis also revealed a significant time by distance interaction; $F(2.66, 231.66) = 3.66, p = .017, \eta_p^2 = .181$, which showed that the distance effect at T2 was significantly reduced compared to T1 (Figure 3).

[INSERT FIGURE 3 HERE]

Non-symbolic addition

A 2 x 2 x 3 mixed ANOVA was carried out to determine the effect of time (T1 and T2), congruency (congruent and incongruent trials; congruent trials are those in which the perceptual features of the array correlated with numerosity) and numerical ratio (1:2, 3:5 and 2:3) on Non-symbolic addition accuracy (see Figure 4). The analysis revealed a significant main effect of time; $F(1, 87) = 31.31, p < .001, \eta_p^2 = .27$. Children performed significantly better on the task at T2 (Mean = 65%), compared to T1 (Mean = 56%).

There was also a main effect of congruency; $F(1, 87) = 375.15, p < .001, \eta_p^2 = .812$. Children tended to perform significantly better on congruent trials (Mean = 87%), compared to incongruent trials (Mean = 34%). There was also a main effect of numerical ratio; $F(2, 174) = 6.39, p = .002, \eta_p^2 = .07$. Bonferroni-corrected pairwise comparisons revealed that children's performance on 1:2 trials (Mean = 63%) was significantly better compared to performance on 2:3 trials (Mean = 58%, $p = .003$).

Stability of basic mathematics-related skills

The analysis also revealed a significant time by congruency interaction; $F(1, 87) = 6.70, p = .011, \eta_p^2 = .07$, which showed that children's performance significantly improved on the incongruent trials across the two time points, but their performance remained stable on the congruent trials (see Figure 3). Finally, there was also a significant congruency by ratio interaction; $F(2, 174) = 4.14, p = .018, \eta_p^2 = .045$. This reflected a significant ratio effect for incongruent trials only.

[INSERT FIGURE 4 HERE]

Number line

A 2 x 2 mixed ANOVA was carried out to examine the effect of time (T1 and T2) and numerical scale (1-10 and 1-20) on Number line estimation accuracy. The analysis revealed a significant main effect of time; $F(1, 86) = 43.96, p < .001, \eta_p^2 = .34$. Children's estimations were closer to the target numbers at T2 (Mean = 126.94 pixels), compared to T1 (Mean = 193.61 pixels). There was also a main effect of scale; $F(1, 86) = 30.66, p < .001, \eta_p^2 = .26$. Children's estimations were closer to the target numbers on the 1-20 scale (Mean = 137.65 pixels), compared to the 1-10 scale (Mean = 182.90).

There was a significant time by scale interaction; $F(1, 86) = 7.41, p = .008, \eta_p^2 = .08$, which showed that the improvement in performance over time was much greater on the 1-20 number line than on the 1-10 number line (Figure 5).

[INSERT TABLE 5 HERE]

Discussion

The current study investigated young children's performance on a series of numerical and non-numerical tasks that have been previously identified as important predictors of early formal mathematics skills in the first years of primary school. Although children in Northern Ireland start primary school at an unusually young age, they were able to perform all of these tasks, and their performance could be measured with acceptable to high reliability. A comparison of children's performance on these tasks between the first and second school years also showed that children's skills on all tasks greatly improved (with a more modest improvement on the Order WM task).

Stability of basic mathematics-related skills

When identifying tasks that form the basis of early competence, and can be indicative of children's later mathematics ability, the focus should be on tasks that are not only related to mathematics skills, but also show stable individual differences over time. We can assume that if performance on a task does not show continuity over time within an individual, the task is not indexing a basic skill. Rather, this pattern indicates that children are in the process of developing new strategies and skills that are relevant for performance on the task. By contrast, tasks that show stability over time *and* are also diagnostic of mathematics skills can be considered more reliable indicators of individual differences in some relevant skills. This distinction is important, because it is often assumed in the mathematics cognition literature that numerical skills originate in some basic abilities that are present from early development.

The results of the current study suggest that, among the tasks which have been previously identified as important predictors of early formal maths skills, individual differences in the order-processing and intelligence measures remained highly consistent between the first two years of primary school. Counting skills and Non-symbolic addition accuracy showed moderate consistency over time, whereas performance on the Number comparison task, and the Number line task showed little consistency.

Regarding the predictive value of these tasks, three tasks showed consistent relationships with counting ability: Order WM, the Daily events task and Number comparison. Based on the results regarding the stability of individual differences on these tasks, we can argue that whereas the two ordering tasks are reliable indicators of individual differences in some early-developing skills related to counting, Number comparison performance can be considered merely diagnostic of the current developmental stage of some relevant skills, as it shows little consistency over time. Indeed, our results suggest that there are important changes in how children perform this task. In particular, children's responses at

Stability of basic mathematics-related skills

T2 were less affected by the numerical distance between the numbers that they had to compare, relative to their performance at T1.

Other tasks that were less stable over time included the Non-symbolic addition and the Number line tasks. On the Non-symbolic addition task, children's performance showed particular improvements in the case of incongruent trials, indicating that they were more able to focus on the number of items presented instead of the perceptual properties of the display (i.e., overall size and surface area), although, as a group, children continued to perform poorly on incongruent trials even at T2. In the case of the Number line task, children showed particular improvement in the 1-20 number line, suggesting that they became increasingly familiar with larger numbers.

Overall, these findings are supportive of an important role for non-numerical ordering skills in the early development of formal mathematics abilities during the first years of primary school, including the ordering of both familiar and novel sequences (e.g., Attout et al., 2014; O'Connor et al., 2018). The results of the current study also extend earlier findings regarding the importance of these skills by showing evidence of the stability of children's temporal ordering skills, as well as their Order WM during the early years of primary school (see also Attout et al., 2014).

Although non-symbolic numerical skills have been cited as being important in early numerical development (e.g., Chen & Li, 2014; Fazio et al., 2014; Schneider et al., 2017), the results of the current study fail to support this assertion, therefore calling into question the claim that the ANS plays a pivotal role in early symbolic number knowledge acquisition (e.g. Chen & Li, 2014; Halberda, Mazzocco & Feigenson, 2008; Holloway & Ansari, 2008; Piazza et al., 2010; Wong, Ho & Tang, 2016). At the same time, the current results are in line with earlier findings that young children struggle to disregard the perceptual properties of non-symbolic displays (Rousselle, Palmers & Noël, 2004; Soltész, Szűcs & Szűcs, 2010), as

Stability of basic mathematics-related skills

although their performance on incongruent trials (in which perceptual properties have to be inhibited in order to respond correctly) improved between T1 and T2, they were still performing poorly on these trials.

It could be argued that the counting task only assessed a small facet of early formal mathematical skills, and because of the forward and backward counting subtasks, it is unsurprising that performance on this task was related to ordering abilities. Nevertheless, O'Connor et al. (2018) found, using data from the same dataset as was used in the current study, that performance on the ordering tasks strongly related to formal mathematics skills (as measured by comprehensive, curriculum-based tests). This result was found both cross-sectionally and longitudinally, and these relations were also somewhat stronger than the correlations between Number comparison and formal mathematics skills, which is similar to the current results.

It is important to note that our results do not question the importance of magnitude processing skills in mathematics development. Indeed, these skills appear to co-develop with other formal mathematics skills, and might be highly diagnostic of children's mathematics competence. Nevertheless, these skills go through some qualitative changes in the first years of school, and show little intra-individual stability, which makes it unlikely that they could form the basis of the development of other mathematical skills. That is, it seems unlikely that these skills could play such a role until performance on these tasks reaches a certain level of stability.

The current findings regarding the low intertemporal stability of non-symbolic magnitude skills are in line with Inglis and Gilmore (2014) who found moderate correlations (around .5) between children's performance on a non-symbolic magnitude comparison task and their performance on the same task one week later. Thus, even within a short time period, children's performance on the task fluctuated considerably. Apart from more general issues

with the reliability of non-symbolic magnitude processing tasks (Gilmore et al., 2013; Inglis & Gilmore, 2013, 2014; Maloney et al., 2010; Price et al., 2012), it has been suggested that performance on these tasks in the case of young children depends on some domain-general skills, such as inhibition (Fuhs & McNeil, 2013) and working memory (Xenidou-Dervou, Van Lieshout & Van der Schoot, 2013). These findings are in line with our proposal that children's performance might reflect a variety of skills and strategies rather than a single underlying ability. These skills and strategies might develop at different rates in different individuals, leading to low consistency in individual differences in overall task performance.

As a limitation of our study, we should note that our results are based on correlations and, as such, do not allow us to establish whether the skills that we assessed played a causal role in the development of mathematical skills. Further research could involve specific interventions to improve some of these skills, and check if this leads to improvements in mathematical ability. Future studies could also investigate the stability of these skills over a longer period of time by tracking children throughout the primary school years.

In summary, the current results suggests that some non-numerical abilities are powerful and reliable predictors of early formal numerical skills. In particular, we found evidence for the important role of ordering skills (see also O'Connor et al., 2018). Nevertheless, future studies could also consider the role of some other non-numerical skills which develop at an early age.

References

- Attout, L., & Majerus, S. (2018). Serial order working memory and numerical ordinal processing share common processes and predict arithmetic abilities. *British Journal of Developmental Psychology*, 36, 285-298.
- Attout, L. & Majerus, S. (2015). Working memory deficits in developmental dyscalculia: The importance of serial order. *Child Neuropsychology*, 21, 432-450.
- Attout, L., Noël, M. P. & Majerus, S. (2014). The relationship between working memory for serial order and numerical development: A longitudinal study. *Developmental Psychology*, 50, 1667-1679.
- Barth, H., Beckmann, L., & Spelke, E. S. (2008). Nonsymbolic, approximate arithmetic in children: Abstract addition prior to instruction. *Developmental Psychology*, 44, 1466.
- Barth, H., La Mont, K., Lipton, J. & Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 14116-14121.
- Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S. & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology*, 46, 545-551.
- Bonato, M., Zorzi, M., & Umiltà, C. (2012). When time is space: evidence for a mental time line. *Neuroscience & Biobehavioral Reviews*, 36, 2257-2273.

- Chen, Q. & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica*, 148, 163-172.
- Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35, 13-21.
- 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, 48-55.
- Eurydice at NFER – National Foundation for Educational Research: Compulsory age of starting school in European countries (2012). Retrieved 3rd January 2019 from <https://www.nfer.ac.uk/nfer/eurydice/publications/compulsory-ageof-starting-school.cfm>
- Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology*, 123, 53-72.
- Friedman, W. J. (1977). The development of children's knowledge of cyclic aspects of time. *Child Development*. 48, 1593-1599.

Stability of basic mathematics-related skills

Friedman, W. J. (1990). Children's representations of the pattern of daily activities. *Child Development, 61*, 1399-1412.

Fuhs, M. W., & McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: Contributions of inhibitory control. *Developmental Science, 16*, 136-148.

Geary, D. C., Brown, S. C., & Samaranayake, V. A. (1991). Cognitive addition: A short longitudinal study of strategy choice and speed-of-processing differences in normal and mathematically disabled children. *Developmental Psychology, 27*, 787.

Gelman, R. & Gallistel, C. R. (1978). *The child's understanding of number*. Washington, DC: American Psychological Association.

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

Gilmore, C., Attridge, N., De Smedt, B., & Inglis, M. (2014). Measuring the approximate number system in children: exploring the relationships among different tasks. *Learning and Individual Differences, 29*, 50-58.

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition, 115*, 394-406.

Stability of basic mathematics-related skills

Halberda, J., Mazocco, M. M. & Feigenson, L. (2008). Individual differences in nonverbal number acuity predict maths achievement. *Nature*, 455, 665-668.

Holloway, I. D. & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: the distance effect and children's mathematical competence. *Journal of Experimental Child Psychology*, 103, 17-29.

Inglis, M. & Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychologica*, 145, 147-155.

Inglis, M., & Gilmore, C. (2013). Sampling from the mental number line: How are approximate number system representations formed?. *Cognition*, 129, 63-69.

Kaufmann, L., Vogel, S. E., Starke, M., Kremser, C., & Schocke, M. (2009). Numerical and non-numerical ordinality processing in children with and without developmental dyscalculia: Evidence from fMRI. *Cognitive Development*, 24, 486-494.

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, 124, 83.

Li, Y., Zhang, M., Chen, Y., Zhu, X., Deng, Z., & Yan, S. (2017). Children's Non-symbolic, Symbolic Addition and Their Mapping Capacity at 4–7 Years Old. *Frontiers in Psychology*, 8, 1203.

Stability of basic mathematics-related skills

Link, T., Huber, S., Nuerk, H. C., & Moeller, K. (2014). Unbounding the mental number line—new evidence on children's spatial representation of numbers. *Frontiers in Psychology, 4*, 1-11.

Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L. & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1–6. *Developmental Science, 1*-13.

Majerus, S., Poncelet, M., Greffe, C., & Van der Linden, M. (2006). Relations between vocabulary development and verbal short-term memory: The relative importance of short-term memory for serial order and item information. *Journal of Experimental Child Psychology, 93*, 95-119.

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*, 154-161.

Morsanyi, K., O'Mahony, E., & McCormack, T. (2017). Number comparison and number ordering as predictors of arithmetic performance in adults: Exploring the link between the two skills, and investigating the question of domain-specificity. *The Quarterly Journal of Experimental Psychology, 70*, 2497-2517.

Morsanyi, K., van Bers, B. M., McCormack, T., & McGourty, J. (2018). The prevalence of specific learning disorder in mathematics and comorbidity with other developmental disorders in primary school-age children. *British Journal of Psychology, 109*, 917-940.

Stability of basic mathematics-related skills

- Morsanyi, K., van Bers, B. M., O'Connor, P. A., & McCormack, T. (2018). Developmental Dyscalculia is Characterized by Order Processing Deficits: Evidence from Numerical and Non-Numerical Ordering Tasks. *Developmental Neuropsychology*, DOI: 10.1080/87565641.2018.1502294
- Moyer, R. S. & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, 215, 1519–1520.
- O'Connor, P. A., Morsanyi, K., & McCormack, T. (2018). Young children's non-numerical ordering ability at the start of formal education longitudinally predicts their symbolic number skills and academic achievement in maths. *Developmental science*, e12645.
- Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D....& Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, 116, 33-41.
- Price, G. R., Palmer, D., Battista, C., & Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: Reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychologica*, 140, 50-57.
- Reeve, R., Reynolds, F., Humberstone, J., & Butterworth, B. (2012). Stability and change in markers of core numerical competencies. *Journal of Experimental Psychology: General*, 141, 649.

Stability of basic mathematics-related skills

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015).

Intelligence and school grades: A meta-analysis. *Intelligence*, 53, 118-137.

Rousselle, L., Palmers, E., & Noël, M. P. (2004). Magnitude comparison in preschoolers:

What counts? Influence of perceptual variables. *Journal of Experimental Child Psychology*, 87, 57-84.

Sasanguie, D., De Smedt, B., & Reynvoet, B. (2017). Evidence for distinct magnitude

systems for symbolic and non-symbolic number. *Psychological Research*, 81, 231-242.

Sasanguie, D., & Vos, H. (2018). About why there is a shift from magnitude to ordinal

processing in the association with arithmetic between first and second grade.

Developmental Science, e12653. doi: 10.1111/desc.12653

Sattler, J. M., & Dumont, R. (2004). Assessment of children: WISC-IV and WPPSI-III

supplement. San Diego: Jerome M. Sattler, Publisher.

Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S., Stricker, J. & De Smedt, B.

(2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Developmental Science*, 20(3), e12372.

Stability of basic mathematics-related skills

- Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., & Luwel, K. (2018). Associations of Number Line Estimation With Mathematical Competence: A Meta-analysis. *Child Development*. DOI: 10.1111/cdev.13068 [Epub ahead of print].
- Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, 75, 428-444.
- Siegler, R. S. & Opfer, J. E. (2003). The development of numerical estimation evidence for multiple representations of numerical quantity. *Psychological Science*, 14, 237-250.
- Soltész, F., Szűcs, D., & Szűcs, L. (2010). Relationships between magnitude representation, counting and memory in 4-to 7-year-old children: A developmental study. *Behavioral and Brain Functions*, 6(1), 13.
- Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, 35, 401-426.
- Vos, H., Sasanguie, D., Gevers, W., & Reynvoet, B. (2017). The role of general and number-specific order processing in adults' arithmetic performance. *Journal of Cognitive Psychology*, 1-14.
- Wechsler, D. (2003) *Wechsler Preschool and Primary Scale of Intelligence – Third UK Edition (WPPSI-III)*. Oxford, England. Psychological Corp.

Stability of basic mathematics-related skills

Wong, T. T. Y., Ho, C. S. H., & Tang, J. (2016). The relation between ANS and symbolic arithmetic skills: The mediating role of number-numerosity mappings. *Contemporary Educational Psychology, 46*, 208-217.

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

Xenidou-Dervou, I., Molenaar, D., Ansari, D., van der Schoot, M., & van Lieshout, E. C. (2017). Non-symbolic and symbolic magnitude comparison skills as longitudinal predictors of mathematical achievement. *Learning and Instruction, 50*, 1-13.

List of figures and tables

Table 1. Table showing the results of three longitudinal studies regarding the stability of cardinal and ordinal measures

Table 2. Descriptive statistics for all measures at T1 and T2 along with the results of paired sample t tests comparing T1 and T2 performance.

Table 3. Correlations between counting skills and all measures at T1 and T2

Figure 1. Bootstrap correlation analysis for task accuracy between T1 and T2

Figure 2. Graph plotting accuracy on the Daily events task by time and trial type (error bars represent the standard error of the mean).

Dotted line – Daily events accuracy at T1

Solid line – Daily events accuracy at T2

Figure 3. Graph plotting accuracy on the Number comparison task by numerical distance and time (error bars represent the standard error of the mean)

Dotted line – Number comparison accuracy at T1

Solid line – Number comparison accuracy at T2

Figure 4. Graph plotting accuracy on the Non-symbolic addition task by ratio and congruency at T1 (left) and at T2 (right). (Error bars represent the standard error of the mean).

Dotted line – Non-symbolic addition accuracy at T1

Solid line – Non-symbolic addition accuracy at T2

Figure 5. Graph plotting mean estimation error on the number line task by time and scale (error bars represent the standard error of the mean)

Dotted line – Number line mean estimation error at T1

Solid line – Number line mean estimation error at T2

