UvA-DARE (Digital Academic Repository)

Four and twenty blackbirds

How transcoding ability mediates the relationship between visuospatial working memory and math in a language with inversion

van der Ven, S.H.G.; Klaiber, J.D.; van der Maas, H.L.J.

Four and twenty blackbirds: how transcoding ability mediates the relationship between visuospatial working memory and math in a language with inversion

Sanne H.G. van der Ven, Jonathan D. Klaiber & Han L.J. van der Maas

Routledge
Taylor & Francis Group

Four and twenty blackbirds: how transcoding ability mediates the relationship between visuospatial working memory and math in a language with inversion
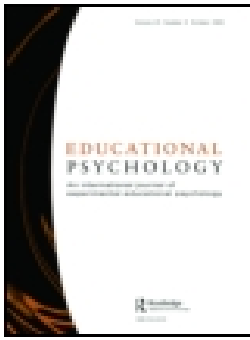
Sanne H.G. van der Ven[a,b], Jonathan D. Klaiber[b] and Han L.J. van der Maas[b]

[a]Cognitive and Motor Development, Educational and Learning Sciences, Utrecht University, Utrecht, The Netherlands; [b]Department of Psychology, Psychological Methods, University of Amsterdam, Amsterdam, The Netherlands

ABSTRACT

Writing down spoken number words (transcoding) is an ability that is predictive of math performance and related to working memory ability. We analysed these relationships in a large sample of over 25,000 children, from kindergarten to the end of primary school, who solved transcoding items with a computer adaptive system. Furthermore, we investigated the nature of transcoding difficulty of over 300 two- and three-digit numbers. All data come from a Dutch sample, meaning that transcoding is complicated by decade-unit inversion: 24 is pronounced as 'four-and-twenty'. Omission to invert the digits of a spoken number when writing it down is an inversion error: the incidence of these declined but did not disappear in later elementary school. Furthermore, transcoding ability mediated the relationship between visuospatial working memory and mathematics performance, a strong effect that declined with age. Inversion error making mediated this same relationship in an inverted U-shaped curve, peaking around grade 2 (8 years old). At the item level, structural characteristics related to inversion and irregular pronunciation of units and decades explained a large part of the variance in item difficulty. We conclude that number transcoding is an important ability to develop mathematical proficiency and discuss the implications of these findings.

One of the most basic numerical skills is writing down number words in symbolic notation: for example, hearing the word 'two' and then writing down the symbol '2'. This ability is often referred to as transcoding. Many studies in this field thus far targeted differences in transcoding ability between children: either cross-cultural comparisons of children speaking languages differing in their verbal number system, or analyses in which transcoding accuracy is predicted from other cognitive abilities, such as working memory. With the present paper we add a comprehensive mediation analysis of transcoding ability as a mediator of the relation between working memory ability and math ability in all primary school grades. In addition, we add a new type of analysis: differences in transcoding difficulty between different number words in the same language, namely Dutch. We investigated characteristics related to the numerical structure and pronunciation of different numbers.

CONTACT Sanne H.G. van der Ven ✉ s.vanderven@uu.nl

Learning transcoding

Number transcoding is related to mathematical performance: compared to normally developing children, arithmetically disabled children made more transcoding mistakes (Moura et al., 2013) and were slower (van Loosbroek, Dirkx, Hulstijn, & Janssen, 2009), while children with good transcoding skills obtained higher math grades (Imbo, Vanden Bulcke, & Fias, 2014).

Transcoding consists of two types of rules: lexical rules, or the correspondence between number words and the numerical symbols 0–9, and syntactic rules, governing the formation of multidigit numbers (Power & Dal Martello, 1990). Learning these rules is a protracted process; transcoding errors in grades 1 and 2 are common (Camos, 2008; Power & Dal Martello, 1990; Seron & Fayol, 1994; Zuber, Pixner, Moeller, & Nuerk, 2009). Mistakes are frequent, especially violations against the syntactic rules, e.g. writing 1003 for 103, or 6100 for 600, in numbers with three or more digits (Camos, 2008; Imbo et al., 2014; Power & Dal Martello, 1990; Zuber et al., 2009).

It has been postulated that accessing the quantity of the number is not necessary while transcoding; instead, a sequence of syntactic rules is applied (Barrouillet, Camos, Perruchet, & Seron, 2004). During development these rules are acquired sequentially. With development, multidigit chunks are formed that are directly retrieved without the need to apply rules. This model, named A Developmental, Asemantic and Procedural Model (ADAPT) was confirmed by simulations that matched children's error patterns (Barrouillet et al., 2004). The authors found that second graders were already able to retrieve two-digit numbers as chunks.

In some languages the order of pronunciation of number words does not match the left-to-right notation order. This phenomenon is illustrated in the following nursery rhyme originating in the eighteenth century: 'Sing a song of sixpence, a pocket full of rye. *Four and twenty* blackbirds, baked in a pie. When the pie was opened, the birds began to sing. Wasn't that a dainty dish, to set before the king?' In modern English this decade-unit inversion property is only remnant in the teen numbers (e.g. 'sixteen'), but it is still present in many other languages: for example, the Germanic languages Dutch, German and Danish, but also Arabic. The native speakers of these four languages together account for hundreds of millions of speakers.

Adults speaking a language with inversion hardly make mistakes in two- and three-digit number transcoding (Zuber et al., 2009). However, for children inversion is difficult. In grade 2, children speaking a language with inversion (Dutch or German) made more transcoding errors than children speaking French, a language without inversion (Krinzinger et al., 2011; but see Imbo et al., 2014). In Czech, which has a system with and one without inversion, the same children made more errors and especially more inversion errors in the system with inversion (Pixner et al., 2011). Another study showed that half of the transcoding errors made by German first graders were inversion errors (Zuber et al., 2009), while these were hardly made by children speaking a language without inversion (Camos, 2008; Power & Dal Martello, 1990). Inversion even seems to affect mathematical performance: Austrian children speaking German (with inversion) slowed down more in addition problems requiring a carry (i.e. when the sum of the units is larger than 10, e.g. 17 + 9) than children speaking Italian (no inversion; Göbel, Moeller, Pixner, Kaufmann, & Nuerk, 2014). Finally, inversion error making in grade 1 predicted addition performance two years later, especially on problems involving a carry (Moeller, Pixner, Zuber, Kaufmann, & Nuerk, 2011).

Transcoding, inversion and working memory

Studies on transcoding accuracy, such as cited in the previous section, focused on children around seven years old. However, transcoding errors may remain in older children under specific circumstances, due to working memory limitations. The ADAPT model (Barrouillet et al., 2004) predicts a relationship between working memory and transcoding performance. Working memory, according to the often-used Baddeley model (Baddeley, 1992; Baddeley & Hitch, 1974), is a system with limited capacity, especially in children (Alloway, Gathercole, & Pickering, 2006; Huizinga, Dolan, & van der Molen, 2006; van der Ven, Kroesbergen, Boom, & Leseman, 2012) but also in adults (Baddeley, 2007; Baddeley & Hitch, 1974). The model consists of two simple storage systems responsible for the storage of verbal (phonological loop) and visuospatial (visuospatial sketchpad) information, and a domain-general attentional control system, the central

executive. The storage systems are usually measured with tasks in which sequences of, respectively, verbal and visuospatial stimuli must be repeated, while central executive tasks involve an additional manipulation, such as the requirement to perform a secondary task (e.g. the to-be-remembered number has to be counted first) or backward span tasks. Later studies confirmed the tri-partite model of working memory, although the distinction between the visuospatial sketchpad and the central executive may not be so clear-cut, especially in children (Ang & Lee, 2008; Gathercole, Pickering, Ambridge, & Wearing, 2004; Kessels, van den Berg, Ruis, & Brands, 2008), because executive involvement is also present in task tapping the visuospatial sketchpad (Ang & Lee, 2008; Fisk & Sharp, 2003; Vandierendonck, Kemps, Fastame, & Szmalec, 2004). In the present study, we therefore use visuospatial working memory when referring to visuospatial working memory tasks while acknowledging that these contain central executive elements.

According to the ADAPT model of transcoding (Barrouillet et al., 2004), each syntactic transcoding rule poses a load on working memory. Under circumstances of low overall task load, the number of mistakes may be very low or absent, as was found by Zuber et al. (2009). When the load is increased, e.g. when the number consists of many digits, the extra load posed by the inversion property may be the final straw to overload working memory. The first aim of the present study was therefore to investigate the degree of inversion error making in older children, under circumstances of higher cognitive load.

A second prediction derived from the ADAPT model is a relation between transcoding and working memory, especially the central executive. This relation has been confirmed in children around seven years of age. One study found a relation between transcoding and a complex memory span in French-speaking children (Camos, 2008). Later studies incorporated all three components of Baddeley's model and found evidence for a unique predictive role of the central executive in children speaking languages with and without inversion (Imbo et al., 2014; Pixner et al., 2011; Zuber et al., 2009) and the visuospatial sketchpad in English-speaking children (Simmons, Willis, & Adams, 2012). Only one study found indications for a unique relation with the phonological loop, which predicted variance in low-performing transcoders only (Imbo et al., 2014).

Thus, most studies found relations between transcoding and especially the interrelated central executive and visuospatial sketchpad. These relations were found in languages with and without inversion, but tended to be stronger in languages with inversion. The strongest indication comes from a study in which the same Czech children were tested in both Czech number word systems, one with and one without inversion: the standardised beta coefficient of the central executive was .44 in the system with inversion and .29 without (Pixner et al., 2011). The relative lack of findings concerning the phonological loop seems to reflect that the main difficulty of number transcoding lies in writing down a number in the correct order, not in recalling the verbally presented number word.

How the relation between working memory and transcoding develops in older children is unknown. The second aim of this study is therefore to investigate how the relation between visuospatial working memory and transcoding in general and inversion error making in particular changes with age. In a language with inversion the strength of these relations may decline, as older children retrieve inverted two-digit numbers as single chunks. However, for multidigit numbers working memory may pose a limit to older children's transcoding ability as well, and therefore we do not expect these relations to fully disappear.

Transcoding and inversion error making as mediators for the relation between visuospatial working memory and mathematics

Relations between visuospatial working memory and mathematical ability have often been reported (Bull & Lee, 2014; De Smedt et al., 2009; Friso-Van den Bos, van der Ven, Kroesbergen, & van Luit, 2013; Holmes & Adams, 2006; Rasmussen & Bisanz, 2005; van der Ven, van der Maas, Straatemeier, & Jansen, 2013). The strength of this relation seems to decrease with age (De Smedt et al., 2009; Holmes & Adams, 2006; Rasmussen & Bisanz, 2005; van der Ven et al., 2013). Explanations for this decrease have thus far focused on children's increasing reliance on verbal coding of information and on verbal calculation procedures due to schooling (De Smedt et al., 2009; Rasmussen & Bisanz, 2005). Here we posit another, additional explanation: a reason for the decrease in this relationship may lie in transcoding, which may

affect mathematical accuracy. When children apply verbal strategies, they need to transcode symbolic numbers to their verbal counterpart, and when finished they need to transcode back to numerals.

These steps require working memory resources. In languages with inversion, the working memory demand may be stronger, especially around seven years of age, when children work with two-digit numbers which they have not yet stored as chunks. Since many of the studies reporting a decrease in relation between working memory and mathematics were carried out in children speaking a language with inversion, specifically this inversion property may explain the decrease, which seems to be especially pronounced at this age. We thus predict that one of the mechanisms through which working memory exerts its action on math performance is that limitations in working memory lead to transcoding errors, which lead to mathematical errors.

In conclusion, our third aim is to investigate whether transcoding ability and inversion error making mediate the relation between working memory and mathematical performance. We expect these effects to be the strongest around grades 1 and 2, and when incorporating inversion as a mediator, we expect the relation between visuospatial working memory and mathematics to be no longer decreasing with age.

Explaining transcoding item difficulty

The fourth and final question addressed in this paper is what the nature is of the difficulties that children have with transcoding. Which characteristics of number words determine this difficulty? Apart from the ADAPT model, this question has hardly received any scientific attention. In the present study, we address this question with a different approach. There are various characteristics of numbers and number words, relating to the pronunciation of number words and to the numerical structure, which may lead to differences in transcoding difficulty. We investigated the effect of these characteristics in Dutch, a language with inversion.

We identified two pronunciation characteristics. The first is word length (number of syllables). The word length effect is well known: longer words are more difficult to recall (Baddeley, Thomson, & Buchanan, 1975; Cowan et al., 1992). This may also affect number word transcoding. The second characteristic is irregular pronunciation. In English there are a few irregularly pronounced decades: e.g. 20 ('twenty'; not 'twoty'). Dutch is comparable to English, the postfix '-tig' is added to the single digits, but, as in English, teen numbers are irregular, and 20 is pronounced irregularly: 'twintig', not 'tweetig'. Smaller irregularities are present for 30 (pronounced 'dertig' instead of 'drietig'), 40 ('veertig' instead of 'viertig') and 80 ('tachtig' instead of 'achtig'). It has been claimed that more regular languages, such as East Asian languages based on ancient Chinese, support children in their understanding of the base ten structure (Geary, Bow-Thomas, Liu, & Siegler, 1996; Ng & Rao, 2010), which partially explains why East Asian children systematically outperform Western children in mathematics (Mullis, Martin, Foy, & Arora, 2012; Ng & Rao, 2010). It is therefore imaginable that also within a language irregular numbers are more difficult to transcode. However, it is also possible that irregular decades serve as facilitating cues preventing inversion errors. For example, when hearing 'twee-en-dertig' ('two-and-thirty'; 32), two cues show that the number cannot be 23: 'twee-' cannot refer to the decade 20, which is pronounced as 'twin-', while 'der-' cannot refer to units, which are pronounced as 'drie'. In Dutch, the decades 10, 20 and 30 differ strongly from their corresponding units (1, 2 and 3), while 40 and 80 have slight irregularities. We refer to these numbers as irregular units (1, 2, 3, 4, 8) and decades (10, 20, 30, 40, 80).

We also identified three characteristics relating to numerical structure: problem size, interdigit distance and number type. Problem size, or the value of the number, is a strong predictor of difficulty in many mathematical domains (Ashcraft & Guillaume, 2009; Campbell & Graham, 1985; Imbo, Vandierendonck, & Rosseel, 2007; Parkman, 1972). In one study a problem size effect has also been demonstrated for transcoding: response latency was longer for larger numbers than for smaller numbers, even when the number of digits was identical (van Loosbroek et al., 2009).

The second characteristic related to numerical structure is interdigit distance: the difference between the two digits in a two-digit number: e.g. in the number 27, the interdigit distance is 5.

Austrian, German-speaking children were less accurate than Italian children in positioning numbers with a large interdigit distance on a number line, while there was no difference for numbers with a small interdigit distance (Helmreich et al., 2011). Austrian children may have misinterpreted the numbers, e.g. positioning 27 at the location of 72, which leads to large deviations for items with a large interdigit distance. In the present study, we investigate if interdigit distance also affects transcoding difficulty.

The third numerical structure characteristic is number type: in two-digit numbers these types are tie (e.g. 66), decade (e.g. 60) and inverted number (e.g. 62, pronounced in Dutch as twee-en-zestig: two-and-sixty). We expect inverted numbers to be more difficult because these require an inversion step that is not necessary in decades. Ties require this inversion step too, but omission also leads to a correct answer; moreover, only one number word needs to be activated in working memory. In three-digit numbers there are even more number types. Numbers can be classified as a multiple of hundred, a decade, as containing three different non-zero digits (e.g. 243), a 'mid-zero' (e.g. 204), a 'full tie' (e.g. 222), but also three types of partial ties, in which two digits are equal: a 'hundred-decade tie' (e.g. 224), a 'decade-unit tie' (e.g. 422) or a 'hundred-unit tie' (e.g. 242). We expect three different non-zero digits to be difficult because of inversion. We also expect numbers with a zero in the middle (e.g. 304) to be difficult, since these numbers require the extra procedural step of filling the empty middle position with this zero (Barrouillet et al., 2004). We further expect hundreds, decades and ties to be easier, with the exception that hundred-decade ties and hundred-unit ties may be difficult; because of the inversion property, in hundred-decade ties the notation of the two equal digits is consecutive but their pronunciation is not (442 is pronounced '*four* hundred two and *forty*', while in hundred-unit ties the pronunciation of the two equal digits is consecutive but their notation is not (424 is pronounced '*four* hundred *four* and twenty).

Aims of the present study

The present study is divided in two parts. In the first part, analyses focus on the first three research questions related to child development: we compared data from Dutch children from all years of primary education, ranging from kindergarten to grade 6. First, we investigated the degree to which these children made inversion errors while transcoding number words. Second, we investigated the relation between transcoding and inversion error making and visuospatial working memory at different ages. Third, we investigated (age differences in) the mediating role of transcoding and inversion error in the relationship between working memory and math performance.

The second part is a detailed analysis dealing with the fourth research question: the nature of the difficulty of two- and three-digit transcoding items. With regression analyses, we investigated how well this difficulty can be predicted by potential characteristics described in the previous section: i.e. the pronunciation characteristics word length and irregular pronunciation of units and decades, and structural characteristics which relate to the numerical structure of the item: problem size, interdigit distance and number type.

To answer these questions, we used the transcoding game from the Dutch version (Rekentuin. nl) of the web-based computer adaptive learning environment named Math Garden (Klinkenberg, Straatemeier, & Van der Maas, 2011), in which children can practice mathematics by playing various games, one of which is number transcoding. Computer algorithms adapt the presented problems to the child's ability: problems are administered such that there is an average probability of .75 to solve the problem correctly. This process ensures that problems are not overly easy: a reasonable cognitive load will be present. This increases the ecological validity of the task: in everyday life, writing down numbers often takes place in an environment in which more sources of cognitive load are present, such as mathematical problem solving. A further advantage of the use of Math Garden is the vast sample size: whereas most studies described in the introduction consisted of fewer than 200 participants, we currently use data from over 25,000 children in the entire range of primary education.

Method

All data used in the present study were obtained with Math Garden. Since this is an unconventional way of collecting data, we present more information about Math Garden and its algorithms in Appendix 1. Further information can be found in Klinkenberg et al. (2011) and Maris and Van der Maas (2012) .

Participants

The participants of Math Garden are predominantly children attending schools using Math Garden as part of their education: approximately 5% of all children in the Netherlands have an account. Participating families and schools gave permission for the (anonymous) use of the data for research purposes; the schools accept the responsibility to inform the parents about the research and voluntary participation. Parents are given the opportunity to refuse the use of their children's data; data from these children are not distributed to the researchers.

In the Netherlands, primary school consists of two years of kindergarten and six years of primary education. The data from kindergarten year 1 proved unreliable because of children's limited computer skills. Therefore, for our user analyses we selected the children from the second year of kindergarten (5 and 6 years old) until the sixth and final year of primary education (11 and 12 years old) who had played at least 30 problems of the transcoding game between February 2013, when the game was introduced, and May 2013. Some children had been registered with highly improbable age-grade combinations (e.g. a 14-year-old in grade 2). Chances are high that for these children mistakes have been made. Therefore, we excluded data from the children whose reported age deviated more than one year from the trimmed mean age of their grade (the 10% extremes on each end were trimmed to determine this mean age). We further removed data from children with reported dates of birth that were improbably frequent (mainly January 1 of any year). The final sample consisted of 25,620 children: 961 (44% girls) from kindergarten, 3915 (47% girls) from grade 1, 5113 (48% girls) from grade 2, 5080 (49% girls) from grade 3, 4399 (48% girls) from grade 4, 3596 (49% girls) from grade 5 and 2556 (46% girls) from grade 6.

Instruments

Math Garden

Math Garden is web-based computer interface, with math game plants in a garden. When a child clicks on a plant, the corresponding game starts. By playing the games, children's plants grow: the higher the ability, the bigger the plant. The games analysed in the present paper are displayed in Figure 1.

The left panel shows the transcoding game: in this game, the child clicks on the play button and a number word is spoken, which the child has to enter in numerals in an answer field, using the computer keyboard. A dimensionality analysis for a subset of two-digit numbers is presented in Appendix 2. For research questions 2 and 3, data from two more games were analysed: the addition game (middle panel), in which the child solves addition problems in a multiple choice format, and the mole game (right panel), a visuospatial working memory game in which moles appear sequentially in a grid with molehills. After the sequence, the child is told to indicate the sequence of the moles, either in the same
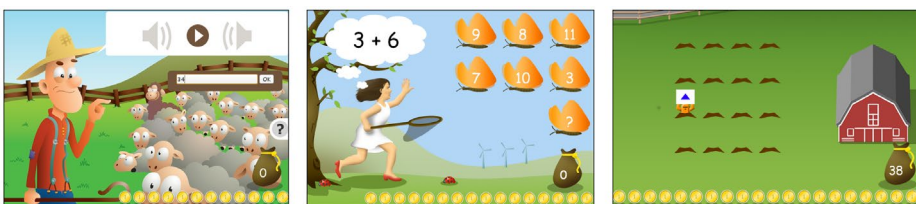

Figure 1. Screenshots of the transcoding game (left), addition game (middle) and mole game (right).

or in reversed order. The difference between same and reversed order has been shown to be only minor; this task taps the visuospatial sketchpad as well as the central executive (van der Ven et al., 2013).

In all games the child is given 20 seconds to answer each item. The remaining time is reflected as a row of coins in the bottom of the screen, from which a coin disappears with each passing second. Upon answering, the correct answer is shown and the child receives the number of remaining coins if the answer was correct, but loses the same number of coins if the answer was incorrect. The rationale of this scoring rule is explained in Appendix 1. The question mark can be clicked if the child does not know the answer: in this case, and when the child did not provide an answer within the time limit, no coins are won or lost and the next item appears. The coins can be used to buy virtual prizes in a trophy cabinet. The child is thus motivated to answer quickly if (s)he knows the answer, but to refrain from answering otherwise. A game ends after 15 items, but children can quit earlier or play the game (with different items) several times.

Each instance of a child answering a problem is used to update both the estimated ability of the child and the estimated difficulty of the item. This enables analyses at two levels: user analyses focus on the (development of) user ability, and item analyses focus on item characteristics.

Math Garden shows a high criterion validity: correlations of r = .78–.84 (Klinkenberg et al., 2011), with norm-referenced math tests (Janssen, Scheltens, & Kraemer, 2005).

Transcoding items

For the second part of the results, the analyses were performed on the difficulty ratings of the items. This means that the population consists of the items, and results can thus only be explained by means of item characteristics (although the children indirectly play a role, since item ratings were determined by the children playing the transcoding game, with the algorithms described in Appendix 1).

The ratings of the two- and three-digit items were extracted from the system as they were in May 2013: the item set contained all two-digit items and 290 three-digit numbers. For two-digit and three-digit numbers separately, we investigated how well the item difficulty ratings could be predicted with the characteristics described in the introduction: first the pronunciation characteristics, then the numerical structure characteristics, and then both types of characteristics together. Finally, an error analysis was carried out on the three-digit items, to see if decade-unit inversions occurred more often than other inversions.

Results

Child development in transcoding and inversion error making

During the four months of the study, all children together solved 4,218,372 problems. Figure 2 shows the distribution of the children's ability ratings by grade: the further towards the right, the higher the rating.

Since as in all latent variable models the scale of the horizontal axis is arbitrary, scores by themselves are not meaningful. Example items illustrate what type of numbers children are capable of transcoding. That is, when the ratings of the child and the problem are equal, the child has a .50 probability of transcoding the number correctly. Because this probability correct is set higher in Math Garden, typically .75, the items presented in Math Garden are also easier: with the setting of .75 correct, item ratings are on average 1.1 points lower than child ratings.

For instance, the most common ability rating of children in grade 4 is around 5, and children performing at this level have a .50 probability of transcoding the number 70312 correctly. When children play Math Garden with the default medium difficulty setting of a probability of .75, the problems presented to them have a difficulty of on average 1.1 point lower than their own ability. If a child chooses to solve easy problems (90% correct), the problem difficulty is on average 2.2 points lower than the child's ability, and if the difficult setting (60%) is chosen, the problem difficulty is 0.4 points lower.
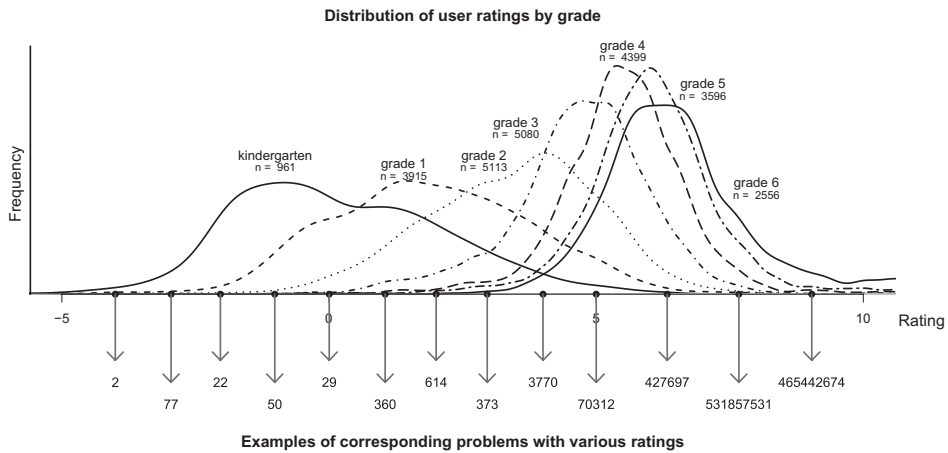
Figure 2. Problem difficulty and child ability estimates by grade. The horizontal axis displays child ability and problem difficulty: both on the same scale. The further towards the right, the more able the child and the more difficult the problem.
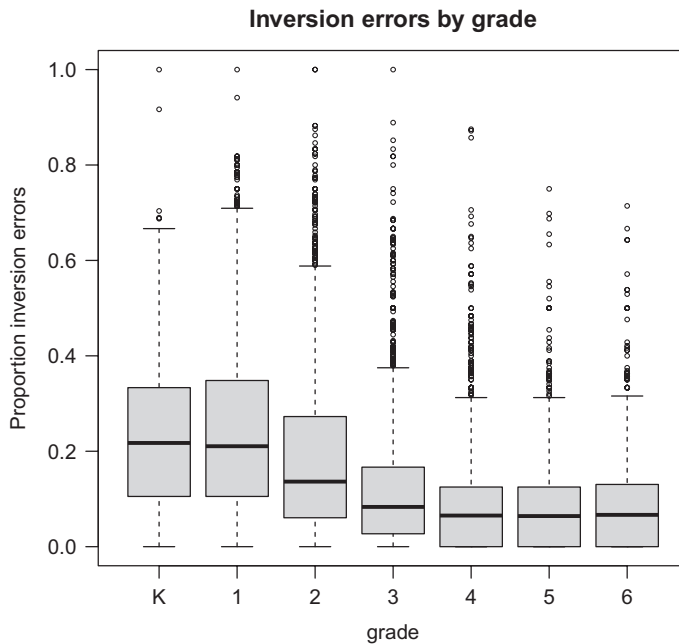
Figure 3. The distribution of the proportion of inversion errors in different grades.

Question 1: prevalence of inversion errors in different grades

In the subset of 12,960 children that had transcoded at least 11 numbers containing at least one inversion, we investigated the incidence of inversion errors. The boxplots in Figure 3 show that inversion errors were frequent in kindergarten and grade 1, and then decreased to a stable average level of around .10 in grades 4–6. A one way-ANOVA showed a significant large effect of grade, $F(6; 12,953) = 421$, $p < .001$, $\eta^2 = .163$. Tukey HSD post hoc tests confirmed that all grades differed significantly from each other, with the exception of kindergarten-grade 1, and grades 4–6. Overall transcoding ability, on the other hand, as shown in Figure 2, continued to increase until grade 6 and this effect was very large, $F(6; 12, 953) = 2520$, $p < .001$, $\eta^2 = .539$ and Tukey HSD post hoc tests confirmed that all grades differed

significantly from each other. This shows that the decrease in inversion error rate is not merely a result of a ceiling effect in the task.

In all grades, the number of children making at least 50% inversion errors was limited and stabilised at a very low percentage after grade 4: 8.3% in kindergarten, 10.9% in grade 1, 7.8% in grade 2, 2.7% in grade 3, 0.9% in grade 4, 0.7% in grade 5 and 0.9% in grade 6. Nevertheless, the number of children not making any inversion error at all was also less than half: 9.1% in kindergarten, 6.5% in grade 1, 13.1% in grade 2, 26.8% in grade 3, 34.5% in grade 4, 35.1% in grade 5 and 42.9% in grade 6. This shows that the majority of children understood the principle of inversion but occasionally made errors.

Questions 2 and 3: inversion, mathematics and visuospatial working memory

Two mediation analyses were performed using the R package Mediation (Tingley, Yamamoto, Hirose, Keele, & Imai, 2013). In the first model, it was tested whether the relation between visuospatial working memory and mathematical ability was mediated by transcoding ability. Mathematical ability was defined by addition ability, a good proxy of overall mathematical ability (Klinkenberg et al., 2011). The same analysis has also been done on a composite score of addition, subtraction and multiplication, but this led to the same results. Data were standardised by grade and the analysis was run for each grade separately.

The procedure consists of four steps. The results are summarised in Figure 4. In step 1, path *c*, the unmediated, direct effect of working memory on addition ability was determined in a linear regression analysis. The results are displayed in the upper part of Figure 4; these are significant at the .05 level if the error bars do not cross the dotted line at 0. The next three steps show the first mediation analysis, shown in the lower left part of Figure 4. In step 2, path *a* was estimated in a linear regression: the effect of independent variable (working memory) on the mediator (transcoding ability). The results with standard errors are visualised next to path *a* in Figure 4: results are also significant at the .05 level if the error bars do not cross the dotted line at 0. In step 3, paths *b* and *c'* were estimated simultaneously in a

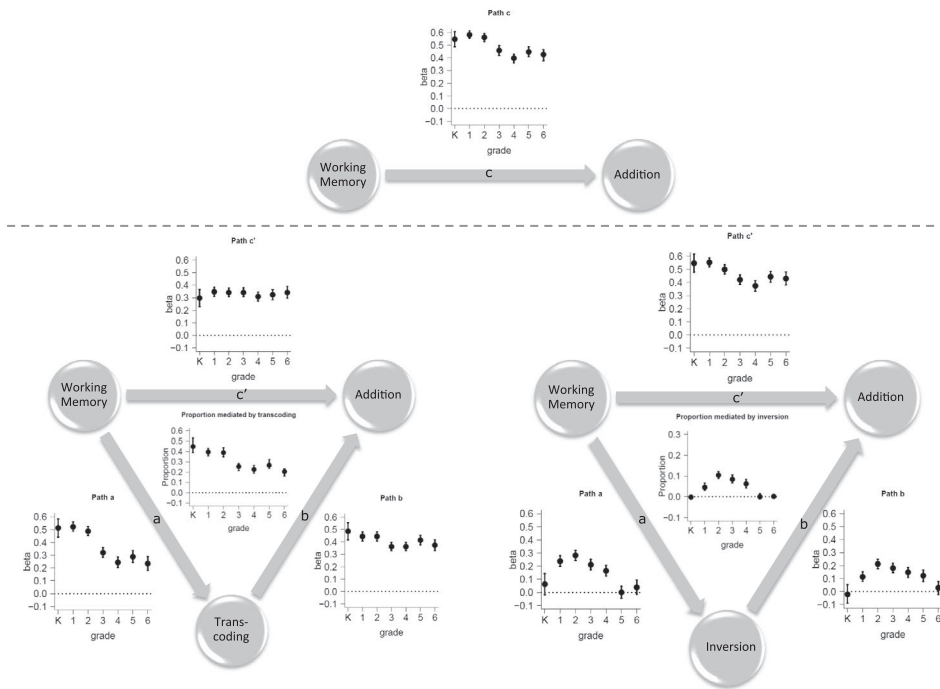**Figure 4.** Results of the mediation analyses by grade: the relation between visuospatial working memory and addition ability, mediated by transcoding ability (upper figure) and inversion accuracy (lower figure). The sizes of the coefficients are presented adjacent to each path of the analysis. The middle graph displays the proportion of the total effect that is explained by the indirect path: $\frac{a \times b}{a \times b + c'}$.

Table 1. Results of t-tests investigating the difference in math ability between the poorest 10% at inversion and all other children.

Grade	df	t	p	Cohen's d
K[a]	573.67	103.21	<.001	0.05
1	2344.11	194.10	<.001	0.36
2	2435.90	135.77	<.001	0.79
3	2414.99	83.63	<.001	0.70
4	2310.76	46.34	<.001	0.55
5	1807.13	15.97	<.001	0.35
6	1298.67	0.27	.784	0.13

[a]K = second year of kindergarten.

second linear regression. These results are displayed in Figure 4 graphically adjacent to the corresponding paths. Note that the total effect of working memory on addition, represented by path c, is equal to $a * b + c'$. Finally, in step 4 a bootstrapping procedure with 50 simulations was used to estimate the proportion of the original path c, explained by the indirect pathway through inversion. This result, with 95% confidence intervals, is displayed in the centre of Figure 4; again, this result is significant at the .05 level if the bars do not cross the horizontal line at 0.

The results show that the direct effect of working memory on addition, and also the effect of transcoding ability on addition, was significant and virtually constant throughout all years of primary education. The relation between working memory and transcoding was also always significant but showed a decline in higher grades. As a consequence, the mediation effect also showed this decline, from .46 in kindergarten to .20 in grade 6.

In the second mediation analysis, shown in the lower right part of Figure 4, the mediator was replaced by inversion ability (proportion of correctly transcribed numbers containing at least one inversion). The results show that a reasonably stable direct path from working memory to math in size with a small decline in higher grades, but both pathways in the indirect path and the proportion mediated by the indirect path show an inverted U-shaped curve, being non-significant in kindergarten and at the end of primary school and significant in the middle, peaking around grade 2.

Since the boxplots in Figure 3 suggest that in each grade a minority of children made inversion errors very frequently, we added a series of Welch two-sample t-tests investigating the difference in math ability between the highest 10 percentiles in inversion error making and the other children in every grade. The results, presented in Table 1, show similar results to the mediation analysis: the difference in math ability was significant but with a very small effect size in kindergarten, showed a peak in grade 2, then declined to non-significance in grade 6.

Question 4: the transcoding difficulty of two- and three-digit numbers

Regression analyses. Two-digit numbers: While Figure 2 displayed a few example items, Figure 5 shows the difficulty of all two-digit numbers between 10 and 99. A series of regression analyses (see Table 2) was performed to explain these difficulties. In the first analysis the pronunciation characteristics were entered, $F(8, 81) = 5.53$, p < .001. Note that each irregular unit was entered separately whereas decades were combined in one variable (except for highly irregular 10). This was done because preliminary analyses indicated that separating the units improved model fit but separating the decades did not. The second analysis was performed on the structural characteristics, $F(4, 85) = 27.20$, p < .001, $R^2 = .56$. In the third analysis, all predictors were entered together in one analysis, $F(12, 77) = 19.26$, p < .001, $R^2 = .56$, which fitted significantly better than both the phonological model, $F(4, 77) = 30.56$, p < .001, and the structural model, $F(8, 77) = 7.27$, p < . 001, showing that both phonological and structural aspects affected item difficulty. All structural variables were significant: items with an inversion were significantly more difficult, and ties were significantly easier than decades. Furthermore, larger numbers were more difficult, while numbers with a large interdigit distance and numbers with a 1 or 2 in the units or 10 as a decade were easier. The predicted values by the final model are presented in Figure 5 as a dotted line.
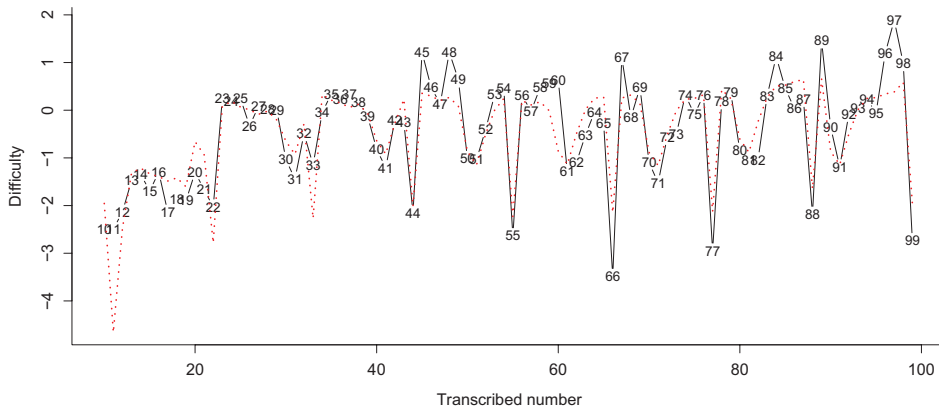
Figure 5. Observed (line) and predicted (dotted) problem difficulties of all two-digit numbers.

Three-digit numbers: A similar series of regression analyses was performed on the 290 three-digit numbers in the transcoding game (see Table 2). Again first the pronunciation characteristics were entered with the same variables as in the two-digit analysis, $F(7, 282) = 21.39$, $p < .001$, $R^2 = .35$. Then the structural analysis was performed, which contained more variables than the two-digit analysis, because the extra digit allowed for more different types, as described in the introduction. The reference category is formed by the numbers consisting of three different non-zero digits (such as 135); thus, a negative regression coefficient means that a type of number was easier than this reference category and a positive regression coefficient means that it was more difficult. The structural characteristics predicted again a higher proportion of variance in item difficulty than the phonological characteristics, $F(8, 281) = 77.48$, $p < .001$, $R^2 = .69$.

In the final regression analysis all predictors were entered. This model fitted the data significantly better than both the phonological model, $F(8, 274) = 50.15$, $p < .001$, and the structural model, $F(7, 274) = 6.93$, $p < .001$. The results show again that most structural predictors were significant: Hundreds and tens were significantly easier than problems with three different non-zero digits, while numbers with a zero in the middle were more difficult. Concerning ties, the result is differentiated. While the presence of a decade-unit tie made a problem easier, full ties were not significantly different and hundred-decade ties (in which the notation of the two equal digits is consecutive but their pronunciation is not), and hundred-unit ties (in which the pronunciation of the two equal digits is consecutive but their notation is not) were even significantly more difficult than items with three different non-zero digits. A higher problem size was also related to a higher difficulty. In the pronunciation part, as in the two-digit item analyses, now the presence of a 1 or 2 in the units made problems significantly easier, while shorter numbers were also easier.

Both regression analyses thus demonstrate that the variables we included can predict item difficulty very well, and especially inversion-related variables affect item difficulty.

Error analysis. An error analysis was carried out on all 127 three-digit numbers in the problem set that consisted of three different non-zero digits. Together these 127 problems had been solved 275,299 times. A mistake was made in 53,522 instances (19%). Of these mistakes, 11,878 (22%) were order mistakes, in which all digits were identified correctly but written in the wrong order. Of all order mistakes, 11,644 (98%) were decade-unit inversions, hundred-decade inversions occurred 15 times (0.1%), hundred-unit inversion errors were made 68 times (0.6%) and a complete shuffle of the answer (e.g. 584 becomes 845 or 458) happened 151 times (1.3%). A binomial test confirmed that the number of decade-unit inversion errors was significantly higher than chance, p < .001. Especially the low incidence of hundred-

Table 2. Results of multiple regression analyses with difficulty of items as dependent variable.

Two-digit numbers (n = 80)

Predictors	(1) Pronunciation R² = .35				(2) Structural R² = .56				(3) All R² = .75			
	β	SE	t	p	β	SE	t	p	β	SE	t	p
Intercept	**-1.21**	**.57**	**-2.16**	**.047**	-.41	.30	-1.37	.175	-.47	.41	-1.12	.266
Pronunciation												
Syllables	.21	.13	1.87	.066					-.03	.14	-.24	.812
Unit = 1	**-1.05**	**.35**	**-3.08**	**.036**					**-1.11**	**.24**	**-4.66**	**<.001**
Unit = 2	-.51	.35	-1.50	.142					**-.54**	**.22**	**-2.43**	**.018**
Unit = 3	.03	.34	-.08	.939					-.11	.23	-.48	.634
Unit = 4	.22	.34	.65	.515					.06	.23	.25	.803
Unit = 8	.23	.34	.67	.503					.09	.23	.40	.687
Decade = 10	**-.98**	**.45**	**-2.19**	**.032**					**-1.08**	**.39**	**-2.79**	**.007**
Decade = 20/30/40/80	.18	.34	.85	.396					.24	.16	1.44	.153
Structural												
Inversion					**.61**	**.26**	**2.35**	**.021**	**.82**	**.37**	**2.20**	**.031**
Tie					**-2.09**	**.39**	**-5.30**	**<.001**	**-1.67**	**.44**	**-3.80**	**<.001**
Problem size					**.01**	**.003**	**4.66**	**<.001**	**.01**	**.004**	**2.03**	**.047**
Interdigit distance					**-.13**	**.04**	**-3.14**	**.002**	**-.07**	**.03**	**-2.01**	**.049**

Three-digit numbers (n = 290)

Predictors	(1) Pronunciation R² = .35				(2) Structural R² = .69				(3) All R² = .74			
	β	SE	t	p	β	SE	t	p	β	SE	t	p
Intercept	**1.92**	**.05**	**41.59**	**<.001**	**2.25**	**.03**	**72.50**	**<.001**	**2.23**	**.04**	**43.75**	**<.001**
Pronunciation												
# Syllables	**.20**	**.02**	**10.43**	**<.001**					**.11**	**.02**	**4.80**	**<.001**
Unit = 1	.09	.11	.83	.408					**-.20**	**.08**	**-2.67**	**.008**
Unit = 2	.05	.12	.47	.642					**-.21**	**.08**	**-2.54**	**.011**
Unit = 3	.20	.12	1.74	.083					-.05	.08	-.58	.562
Unit = 4	**.37**	**.11**	**3.50**	**<.001**					.02	.08	.69	.490
Unit = 8	**.31**	**.12**	**2.50**	**.013**					-.02	.08	-.27	.791
Decades = 20/30/40/80	**-.13**	**.06**	**-2.20**	**.029**					-.07	.04	-1.83	.068
Structural												
Hundreds (e.g. 300)					**-2.10**	**.12**	**-17.34**	**<.001**	**-1.72**	**.15**	**-11.10**	**<.001**
Tens (e.g. 320)					**-.80**	**.05**	**-16.04**	**<.001**	**-.61**	**.07**	**-7.91**	**<.001**
Midzero (e.g. 306)					.15	.09	1.65	.100	**.44**	**.10**	**4.02**	**<.001**
Full tie (e.g. 888)					-.18	.20	-.90	.370	-.08	.20	-.39	.695
Hundred-decade tie (e.g. 443)					.13	.09	1.37	.173	**.22**	**.09**	**2.36**	**.019**
Decade-unit tie (e.g. 433)					**-.77**	**.09**	**-8.79**	**<.001**	**-.74**	**.08**	**-9.04**	**<.001**
Hundred-unit tie (e.g. 434)					**.23**	**.09**	**2.65**	**.008**	**.19**	**.08**	**2.36**	**.019**
Problem size					<.01	<.01	-1.11	.279	**<.01**	**<.01**	**-2.85**	**.005**

Note: Significant results are printed in boldface.

decade inversions compared to the high incidence of decade-unit inversions suggests that the inversion property of the language is the real cause, as opposed to mere typos.

Discussion

In the present study transcoding in a language with inversion was studied. In the first part, we investigated transcoding ability in children. We investigated the degree of inversion error making, the relation between working memory and transcoding, and the mediating role of transcoding ability in general and inversion error making in particular in the relation between visuospatial working memory and mathematics. In the second part, we investigated which item characteristics affect transcoding difficulty.

The results showed that inversion errors in transcoding were common in the lower grades and stabilised at a low but constant level in grades 4–6. In these higher grades, the curriculum is long past the transcoding of numbers; indeed only few children made more than 50% inversion errors. Nevertheless, most children made occasional inversion errors. Because of the adaptive algorithms, children in higher grades transcribed larger numbers (see also Figure 1). Thus, to some degree inversion errors are persistent when the cognitive load is high enough.

Inversion error making was related to visuospatial working memory in an inverted U-shaped age pattern: peaking in grade 2 when the beta coefficient was around .3, similar to other studies around this age range (Pixner et al., 2011; Zuber et al., 2009). This confirms the statement by Zuber et al. (2009) that the central executive is involved in mastering the principle of inversion. The later decrease in the size of this relation suggests that older children have mastered the syntactical principles of inversion. However, the small remaining relation, as well as the persistence of occasional inversion errors, suggest that two-digit numbers may not always be retrieved from memory as single chunks, as stated previously (Barrouillet et al., 2004).

The relation with visuospatial working memory was not specific to inversion error making: the relation with overall transcoding ability was even stronger, starting with a rather constant beta coefficient around .50 in the lowest three grades, after which it suddenly dropped to around .25 in the highest four grades. The difference in strength might again be due to the syntactic principles of transcoding, of which inversion is a prominent one, that are fully mastered in higher grades. The remaining relation with working memory in these years may be due to the larger size of the presented numbers, meaning that the limits of (verbal) working memory capacity are approached.

The results of the mediation analyses suggest that the nature of inversion errors changes with age. Inversion error making was a non-significant mediator, and a non-significant predictor of math ability in kindergarten. In kindergarten children mainly solve one-digit addition problems, in which inversion does not play a role. From grade 1 onwards, inversion error making predicted math ability and mediated the relation between working memory and math. This confirms the findings by Moeller et al. (2011) that inversion error incidence predicted later math ability. These relations peaked in grade 2, when children start solving larger multidigit problems, and declined in higher grades, to non-significance again in grade 6, suggesting that errors were then most likely small attentional slips. Even the 10% children performing most poorly on the inversion items in grade 6 did not obtain significantly lower scores in math than their peers who made fewer inversion errors. This suggests that inversion errors are persistent and continue to affect math performance quite long in primary school, but its influence declines with age.

The results thus indicate that around grade 2 in primary school, children with poor visuospatial working memory perform more poorly in math, partially because they struggle to master the principles of inversion. This partially explains the previous findings that the relationship between visuospatial working memory and mathematics declines with age (De Smedt et al., 2009; van der Ven et al., 2013) – although contrary to expectations, the decreasing trend with age did not fully disappear. For children struggling with inversion a (temporary) different counting system without inversion may be helpful. A promising intervention study showed that teaching an alternative regular Asian-language-based counting system in which 13 is pronounced as 'one ten three' improved number sense in Dutch children with mild intellectual disabilities (Van Luit & Van der Molen, 2011). The finding by Pixner and colleagues

(2011) that the same Czech children made more transcoding errors in the Czech system with inversion than in the system without, further confirms the merits of this approach. A reform in Norwegian in 1951 has shown that it is even possible to introduce a new number word system without inversions in education (Jahr, 1989).

However, we also found a consistent relation between overall transcoding ability and mathematics. Transcoding ability was a far stronger mediator in the relation between visuospatial working memory and mathematics than inversion, and although the mediating effect decreased with age, now the age trend in the relation between working memory and mathematics fully disappeared. The continuing relation between transcoding and mathematics shows that math results are likely not only influenced by the ability to understand the problem and carry out the correct procedures, but also simply by writing the numbers down correctly. Given the scarcity of research on transcoding in older normally developing children, this is a 'blind spot' in mathematics research.

In the second part, we analysed which number characteristics affect transcoding difficulty, with a comprehensive item analysis. Factors related to number word inversion played a large role. In the two-digit analysis, inverted numbers were more difficult than decades and hundreds. However, this might also be the case in language without inversion, as the ADAPT model predicts fewer steps for these numbers in other languages as well (Barrouillet et al., 2004). Nevertheless, the regression analyses also revealed other effects of inversion: two-digit ties were easier than non-ties, and three-digit ties were only easier if the two digits that were equal were in decade and unit position, and even more difficult than non-ties if they were in hundred and decade or hundred and decade position. Moreover, the error patterns reflected difficulty with inversion: of the five possible order mistakes in three-digit numbers, virtually all (98%) were decade-unit inversions.

Together, these results suggest that transcoding is affected by the linguistic properties of a language, and that the ADAPT model needs to be expanded with an extra inversion rule that must be mastered by speakers of languages with inversion. However, the results show more. The ADAPT model would not predict a difference in difficulty between ties and non-ties: also with inversion included in the model, both number types have an equal number of required steps. The even higher difficulty of some ties suggests that transcoding is not a mere asemantic syntactical sequence of steps, at least not in languages with inversion. Instead it suggests semantic activation leading to a conflict that must be resolved: the two identical digits are perceptually salient, but their locations differ between the auditory input and the symbolic output.

The results showed more evidence in favour of semantic (co-)activation. The effect of problem size was marginally significant in the two-digit numbers and significant in the three-digit numbers. Moreover, a small effect of interdigit distance was found: numbers with a large interdigit distance were easier. This suggests that some activation of the numerical magnitude takes place, which in the case of a large interdigit distance prevents a reversal. At first sight, this seems to contradict the results by Helmreich et al. (2011), who found that German-speaking children showed larger deviations when locating numbers with a larger interdigit distance on a number line, while these numbers were easier in our study. However, Helmreich et al. (2011) argued that these larger deviations were probably due to the fact that the difference between the target number and its inverted number are larger, so if a number is misinterpreted, the deviation is larger. This is not the case in our study: the score is not influenced by the distance between the provided and the correct answer. It is possible that the absolute number of mismatches was also smaller in the Helmreich et al. (2011) study: this may be obscured by the larger deviations in these cases.

The above-mentioned effects favouring semantic processing might change with development. Imbo et al. (2014) showed that in skilled transcoders, transcoding happened predominantly asemantically, i.e. without accessing the value of the number, while less skilled transcoders at least sometimes used a semantic route. In our data, the children were presented with problems with a high but not overly high probability of a correct answer, so they were not overly skilled in the problems they solved.

Of all factors related to the pronunciation of number words, only the presence of 1 or 2 as unit was consistent significant predictors: these items were easier than items with a different unit. The irregularity

of decade pronunciation has been put forward as one of the factors explaining the disadvantage in mathematics in children growing up in Western countries compared to children speaking fully regular East Asian languages (Ng & Rao, 2010). But in a language with the inversion property two negatives make a positive: the irregularity functions as a cue that inversion is necessary. This effect was especially strong in the units and less in the decades, which may be explained by the fact that these come later in pronunciation; apparently, by that time the inversion has already been processed.

Some limitations must be mentioned: the data were collected with an online computer adaptive system, meaning that there is less control over the circumstances of data collection than in a traditional research setting. This method of data collection, however, also enabled a vast sample size, and its ecological validity is large, since the data were collected while children were practicing as part of their daily routine. Another limitation is that the data were gathered in one language only: Dutch. A replication in other languages with other numerical characteristics is desirable.

In conclusion, the results show that inversion affects children's performance to transcribe numbers. While this problem is most prominent in young children, it never fully disappears. This suggests that the increasing load of inversion gives speakers of a language with inversion a small but permanent disadvantage that appears in demanding situations. This disadvantage affects quite a number of speakers: Arabic in its different varieties alone is the fifth largest language in the world. The general effect of transcoding ability, however, which had a relation with mathematics that was stronger and more persistent with age than the effect of inversion alone, suggests that transcoding is more than inversion alone. It might therefore also play an important role in other languages. Furthermore, the difficulty analyses revealed effects of inversion, and effects showing that transcoding may not be fully asemantic. This shows that especially languages with mismatches between pronunciation and notation of number words can reveal effects of semantic processing that would possibly go unnoticed in fully regular languages. Both findings warrant further research, preferably cross-linguistic studies comparing these effects in fully regular languages, various languages with different types of irregularities such as French and Danish, and languages with a two parallel number word systems, with and without inversion, such as Norwegian and Czech.

Acknowledgements

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

References

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development, 77*, 1698–1716. doi:10.1111/j.1467-8624.2006.00968.x

Ang, S. Y. & Lee, K. (2008). Central executive involvement in children's spatial memory. *Memory, 16*, 918–933. doi:10.1080/09658210802365347

Ashcraft, M. H. & Guillaume, M. M. (2009). Mathematical cognition and the problem size effect. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 51, pp. 121–151). Amsterdam: Academic Press.

Baddeley, A. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.

Baddeley, A. D. (1992). Working memory. *Science, 255*, 556–559.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation*, (Vol. 8, pp. 47–89). New York, NY: Academic Press.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575–589. doi:10.1016/S0022-5371(75)80045-4

Barrouillet, P., Camos, V., Perruchet, P., & Seron, X. (2004). ADAPT: A developmental, asemantic, and procedural model for transcoding from verbal to arabic numerals. *Psychological Review, 111*, 368. doi:10.1037/0033-295X.111.2.368

Bates, D. M. (2010). *Mixed-effects modeling with R*. Retrieved from http://www.wordlatex.com/static/templates/pdf/book_B.pdf

Bull, R. & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives, 8*, 36–41. doi:10.1111/cdep.12059

Camos, V. (2008). Low working memory capacity impedes both efficiency and learning of number transcoding in children. *Journal of Experimental Child Psychology, 99*, 37–57. doi:10.1016/j.jecp.2007.06.006

Campbell, J. I. D. & Graham, D. J. (1985). Mental multiplication skill: Structure, process, and acquisition. *Canadian Journal of Psychology/Revue canadienne de psychologie, 39*, 338–366. doi:10.1037/h0080065

Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language, 31*, 1–17. doi:10.1016/0749-596X(92)90002-F

De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., & Ghesquière, P. (2009). Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. *Journal of Experimental Child Psychology, 103*, 186–201. doi:10.1016/j.jecp.2009.01.004

Elo, A. (1978). *The rating of chess players, past and present*. New York, NY: Arco.

Fisk, J. E. & Sharp, C. A. (2003). The role of the executive system in visuo-spatial memory functioning. *Brain and Cognition, 52*, 364–381. doi:10.1016/S0278-2626(03)00183-0

Friso-Van den Bos, I., van der Ven, S. H. G., Kroesbergen, E. H., & van Luit, J. E. H. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review, 10*, 29–44. doi:10.1016/j.edurev.2013.05.003

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*, 177–190. doi:10.1037/0012-1649.40.2.177

Geary, D. C., Bow-Thomas, C. C., Liu, F., & Siegler, R. S. (1996). Development of arithmetical competencies in Chinese and American children: Influence of age, language, and schooling. *Child Development, 67*, 2022–2044. doi:10.2307/1131607

Göbel, S. M., Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H.-C. (2014). Language affects symbolic arithmetic in children: The case of number word inversion. *Journal of Experimental Child Psychology, 119*, 17–25. doi:10.1016/j.jecp.2013.10.001

Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H.-C., & Moeller, K. (2011). Language effects on children's nonverbal number line estimations. *Journal of Cross-Cultural Psychology, 42*, 598–613. doi:10.1177/0022022111406026

Holmes, J. & Adams, J. W. (2006). Working memory and children's mathematical skills: Implications for mathematical development and mathematics curricula. *Educational Psychology, 26*, 339–366. doi:10.1080/01443410500341056

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44*, 2017–2036. doi:10.1016/j.neuropsychologia.2006.01.010

Imbo, I., Vanden Bulcke, C., & Fias, W. (2014). Sixty-four or four-and-sixty? The influence of language and working memory on children's number transcoding. *Frontiers in Psychology, 5*, 1–10. doi:10.3389/fpsyg.2014.00313

Imbo, I., Vandierendonck, A., & Rosseel, Y. (2007). The influence of problem features and individual differences on strategic performance in simple arithmetic. *Memory & Cognition, 35*, 454–463. doi:10.3758/bf03193285

Jahr, E. H. (1989). Language planning and language change. In L. E. Breivik & E. H. Jahr (Eds.), *Trends in linguistics. Studies and monographs* (Vol. 43, pp. 99–113). New York, NY: Mouton de Gruyter.

Janssen, J., Scheltens, F., & Kraemer, J.-M. (2005). *Leerling-en onderwijsvolgsysteem rekenen-wiskunde* [Student and education tracking system arithmetic-mathematics]. *Arnhem*, The Netherlands: Cito.

Kessels, R. P. C., van den Berg, E., Ruis, C., & Brands, A. M. A. (2008). The backward span of the corsi block-tapping task and its association with the WAIS-III digit span. *Assessment, 15*, 426–434. doi:10.1177/1073191108315611

Klinkenberg, S., Straatemeier, M., & Van der Maas, H. L. J. (2011). Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education, 57*, 1813–1824. doi:10.1016/j.compedu.2011.02.003

Krinzinger, H., Gregoire, J., Desoete, A., Kaufmann, L., Nuerk, H.-C., & Willmes, K. (2011). Differential language effects on numerical skills in second grade. *Journal of Cross-Cultural Psychology, 42*, 614–629. doi:10.1177/0022022111406252

Maris, G. & Van der Maas, H. (2012). Speed-accuracy response models: Scoring rules based on response time and accuracy. *Psychometrika, 77*, 615–633. doi:10.1007/s11336-012-9288-y

Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., & Nuerk, H. C. (2011). Early place-value understanding as a precursor for later arithmetic performance – A longitudinal study on numerical development. *Research in Developmental Disabilities, 32*, 1837–1851. doi:10.1016/j.ridd.2011.03.012

Moura, R., Wood, G., Pinheiro-Chagas, P., Lonnemann, J., Krinzinger, H., Willmes, K., & Haase, V. G. (2013). Transcoding abilities in typical and atypical mathematics achievers: The role of working memory and procedural and lexical competencies. *Journal of Experimental Child Psychology, 116*, 707–727. doi:10.1016/j.jecp.2013.07.008

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *The TIMSS 2011 international results in mathematics*. Boston, MA: International Association for the Evaluation of Educational Achievement.

Ng, S. S. N. & Rao, N. (2010). Chinese number words, culture, and mathematics learning. *Review of Educational Research, 80*, 180–206. doi:10.3102/0034654310364764

Parkman, J. M. (1972). Temporal aspects of simple multiplication and comparison. *Journal of Experimental Psychology, 95*, 437–444.

Pixner, S., Zuber, J., Heřmanová, V., Kaufmann, L., Nuerk, H. C., & Moeller, K. (2011). One language, two number-word systems and many problems: Numerical cognition in the Czech language. *Research in Developmental Disabilities, 32*, 2683–2689. doi:10.1016/j.ridd.2011.06.004

Power, R. J. D. & Dal Martello, M. F. (1990). The dictation of Italian numerals. *Language and Cognitive Processes, 5*, 237–254. doi:10.1080/01690969008402106

Rasmussen, C. & Bisanz, J. (2005). Representation and working memory in early arithmetic. *Journal of Experimental Child Psychology, 91*, 137–157. doi:10.1016/j.jecp.2005.01.004

Seron, X. & Fayol, M. (1994). Number transcoding in children: A functional analysis. *British Journal of Developmental Psychology, 12*, 281–300. doi:10.1111/j.2044-835X.1994.tb00635.x

Simmons, F. R., Willis, C., & Adams, A.-M. (2012). Different components of working memory have different relationships with different mathematical skills. *Journal of Experimental Child Psychology, 111*, 139–155. doi:10.1016/j.jecp.2011.08.011

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2013). *Mediation: R package for causal mediation analysis*. Retrieved from http://cran.r-project.org/web/packages/mediation/vignettes/mediation.pdf

van der Ven, S. H. G., Kroesbergen, E. H., Boom, J., & Leseman, P. P. M. (2012). The development of executive functions and early mathematics: A dynamic relationship. *British Journal of Educational Psychology, 82*, 100–119. doi:10.1111/j.2044-8279.2011.02035.x

van der Ven, S. H. G., van der Maas, H. L. J., Straatemeier, M., & Jansen, B. R. J. (2013). Visuospatial working memory and mathematical ability at different ages throughout primary school. *Learning and Individual Differences, 27*, 182–192. doi:10.1016/j.lindif.2013.09.003

van Loosbroek, E., Dirkx, G. S., Hulstijn, W., & Janssen, F. (2009). When the mental number line involves a delay: The writing of numbers by children of different arithmetical abilities. *Journal of Experimental Child Psychology, 102*, 26–39.

Van Luit, J. E. H. & Van der Molen, M. J. (2011). The effectiveness of Korean number naming on insight into numbers in Dutch students with mild intellectual disabilities. *Research in Developmental Disabilities, 32*, 1941–1947. doi:10.1016/j.ridd.2011.03.028

Vandierendonck, A., Kemps, E., Fastame, M. C., & Szmalec, A. (2004). Working memory components of the Corsi blocks task. *British Journal of Psychology, 95*, 57–79. doi:10.1348/000712604322779460

Zuber, J., Pixner, S., Moeller, K., & Nuerk, H.-C. (2009). On the language specificity of basic number processing: Transcoding in a language with inversion and its relation to working memory capacity. *Journal of Experimental Child Psychology, 102*, 60–77. doi:10.1016/j.jecp.2008.04.003

Appendix 1

This appendix consists of two parts: a more detailed explanation of the *Math Garden* computer-adaptive algorithms, and a dimensionality analysis of a subset of items from the transcoding game.

Math Garden computer-adaptive algorithms

The *Math Garden* computer adaptive technology works in an iterative process consisting of four steps: (1) problem selection, (2) determining the expected score, (3) the child solving the problem and (4) updating the child ability and problem difficulty estimates based on the difference between the expected and obtained score.

In step 1, a problem is selected for the child. This selection is based on past performance of both the player and the problems (by other players solving the same problem). This past performance yielded an ability estimate of the player and difficulty estimates of the problems. A problem is selected for which the child has an expected probability of on average .75 (SD = .1) of giving the correct answer within the time limit, according to the Rasch model shown in Equation (1):

$$P(X_{ij} = 1 | \theta_j) = \frac{e^{\theta_j - \beta_i}}{1 + e^{\theta_j - \beta_i}} \tag{1}$$

In this equation P = probability, X = the accuracy (1 = correct, 0 = incorrect) of the answer of player j with ability estimate θ_j on item i with difficulty β_i; the probability increases with increasing player ability and decreases with increasing problem difficulty.

In step 2, the expected score is determined. Math Garden uses a scoring system that incorporates not only accuracy but also speed: the high speed high stakes principle (HSHS; Maris & Van der Maas, 2012). The HSHS principle entails that the score equals the proportion of remaining time for correct answers, and the proportion of remaining time multiplied by −1 if the answer was incorrect. The score is thus largest, positively or negatively, for fast answers and linearly decreases/increases to 0 as the time approaches the limit. The principle is shown in Figure 6. For example, if an immediate answer is given, the score is +1 (correct answer) or −1 (incorrect answer). If an answer is given with 75% of the time remaining, the
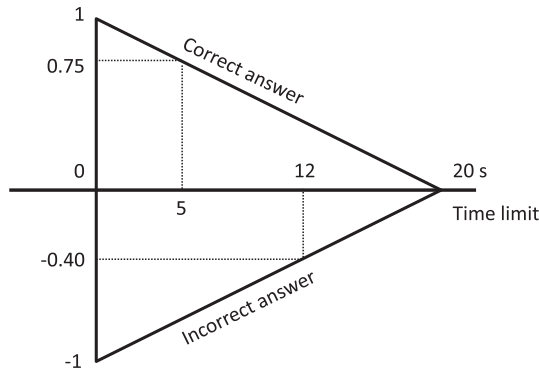
Figure 6. Illustration of the high speed high stake principle. With a time limit of 20 seconds, a correct answer after 5 s yields a score of 0.75; an incorrect answer after 12 s yields a score of −0.40.

score is 0.75 (correct) or −0.75 (incorrect). All values between −1 and 1 are possible. With this rule, fast correct responses are rewarded while fast incorrect guesses are punished.

The child's expected score on the problem is estimated with the extended Rasch model, presented in Maris and Van der Maas (2012). This model is shown in equation (2):

$$E(S_{ij}|\theta_j) = \frac{e^{2(\theta_j - \beta_i)} + 1}{e^{2(\theta_j - \beta_i)} - 1} - \frac{1}{\theta_j - \beta_i} \tag{2}$$

In this equation, E = expected, S = score, θ_j = ability of player j, and β_i = difficulty of problem i. The outcome of Equation (2) is thus the expected score of player j with ability estimate θ_j on item i with difficulty estimate β_i. As Equation (2) shows, this expected score is based on the difference between β_i and θ_j: this expected score ranges from close to −1 (when problem difficulty is far higher than child ability), through 0 (when both are equal), to almost 1 (when child ability is far higher than problem difficulty).

In step 3, the child solves the problem and the actual score is determined, using the HSHS principles as displayed in Figure 1. Then step 4 is executed, a new feature in *Math Garden*: an on-the-fly item calibration system. This means that directly after the problem is solved, the child's obtained score is used to update the estimate of the child's ability and the difficulty of the item. The expected score derived from Equation (2) is compared to the actual score. Then, following a procedure invented by Elo (1978) and also used in e.g. competitive chess to rank players, the child's ability is adjusted upward if the obtained score was higher than expected and downward if the score was lower than expected; the larger the discrepancy between observed and expected score, the larger the adjustment. In a similar fashion, the problem difficulty is adjusted downward if the obtained score was higher than expected (as the problem was easier than expected) and upward if the obtained score was lower than expected. Then the cycle starts again: a new problem is selected, based on the updated estimates. This way, estimates of child ability and problem difficulty are obtained reliably with an on-the-fly algorithm that enables adaptive problem selection such that children only solve problems that are suitable for their level of development.

Appendix 2. The algorithms described above work with an underlying ability estimate. Therefore they only function optimally if there is one underlying ability scale, and all items measure this same ability. In other words, tests based on the 1 PL model such as Math Garden only function optimally if the assumption of unidimensionality is met. This assumption can be tested with a dimensionality analysis. However, these analyses require a full data matrix: a data-set in which every user solved every item. The computer-adaptive nature of Math Garden prohibits this. To enable a dimensionality analysis, we selected a subset of items and users for whom we could obtain a full data matrix: 14,898 Math Garden users (mean age: 8.92 years; SD: 1.49; 50.6% female) each solved 65 items. Fifty-one items contained an inversion: 21, 23, 24, 25, 26, 27, 28, 29, 31, 32, 34, 35, 36, 37, 38, 39, 41, 42, 43, 47, 48, 49, 51, 52, 53, 54, 58, 59, 61, 62, 63, 64, 65, 68, 69, 71, 72, 73, 74, 75, 81, 82, 83, 84, 85, 86, 87, 91, 92, 93, 94 and fourteen items did not contain an inversion: 20, 22, 30, 33, 40, 44, 50, 60, 66, 70, 80, 88, 90 and 100. With a general linear mixed model analysis (Bates, 2010), two models were compared to investigate whether transcoding can be considered unidimensional or bidimensional: two separate transcoding abilities for inverted and non-inverted numbers. The first, unidimensional model was a model with one fixed

effect, whereas the bidimensional model was a model with two fixed effects: one for inverted items and one for non-inverted items. The results indicate that the fixed effect estimates are significant in both models ($p < .001$). The BIC for model comparison was 236,832 for the unidimensional model and 236,818 for the bidimensional model. The bidimensional model is thus favoured, but only slightly. To test the correlation between the two dimensions, a third model was created: a Fixed-Random-Effects model, which was similar to the bidimensional model, with the addition of a random effect to allow the fixed effects to vary between users. This Fixed-Random Effects model revealed a large correlation between the two dimensions, $r = .81$, which means that for users there is not much difference between inverted and non-inverted items. In other words, we may assume one underlying dimension. Moreover, a mix of inversion and non-inversion items in the transcoding game is essential for testing transcoding ability. If there are only (non)-inversion items, users automatically enter a number in the (same) reversed sequence as they hear it. The consequence would be an overestimation of user ability and an underestimation of item difficulty. The mix of inversion and non-inversion items ensures that the transcoding game works well and guarantees a more precise measurement of transcoding ability.