Learning and Individual Differences 27 (2013) 182–192

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Visuospatial working memory and mathematical ability at different ages throughout primary school ☆

Sanne H.G. van der Ven [a,*], Han L.J. van der Maas [a], Marthe Straatemeier [a], Brenda R.J. Jansen [b]

[a] Department of Psychology, Psychological Methods, University of Amsterdam, Weesperplein 4, 1018 XA Amsterdam, The Netherlands
[b] Department of Psychology, Developmental Psychology, University of Amsterdam, Weesperplein 4, 1018 XA Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 16 November 2012
Received in revised form 27 June 2013
Accepted 4 September 2013

Keywords:
Visuospatial
Working memory
Mathematics
Children
Adaptive testing

ABSTRACT

In this study, three aims were addressed: (1) validating a visuospatial working memory task in Math Garden, an adaptive online tool in which item difficulty and person ability are determined on the fly, (2) investigating the contribution of different item characteristics to the difficulty of the visuospatial working memory items, and (3) investigating relations between visuospatial working memory and various math domains at different ages. The method was validated by showing that item ratings were stable and grade differences in ability were significant. Regression analyses on the item level showed that not only sequence length, but also other characteristics, such as type of task (forward or backward), explained variance in item difficulty. Finally, regression analyses on the child level showed that visuospatial working memory and mathematics were significantly related: especially for addition and subtraction in the lower grades. For multiplication and division this relationship was weaker and without age trend.

1. Introduction

In this study, the relationship between visuospatial working memory and mathematics in different math domains and at different ages is examined in a somewhat uncommon design in the scientific field: schools and families buy accounts that enable children to practice various math games online, in school and at home, at any time, while their responses and progress are logged for scientific purposes. In Math Garden, as the system is called (Dutch: Rekentuin), we use computer adaptive technology to adapt the difficulty of the presented problems to the ability level of the child, such that children always play and practice at their own level. This way, learning takes place faster than without adaptive item administration (Holmes, Gathercole, & Dunning, 2009; Klingberg et al., 2005). Indeed, Math Garden has been shown to be an effective tool to improve math skills (Jansen et al., 2013).

This method of data collection differs from traditional measures but it offers several advantages, a number of which are listed below. First, since the presence of an experimenter in the room is not required, sample sizes are far larger than in most other studies: in 2012 there were approximately 50,000 active users, predominantly primary school children. The large number of participating schools offers a representative selection of children: the potential bias in traditional research that only certain types of schools are willing to participate in scientific research, is reduced. Second, the amount of data per participant is also extensive, because they are tested on various domains. Third, because the tests are adaptive and span a wide difficulty range, it is possible to assess the same skill in participants differing greatly in ability (e.g., kindergartners and adults), and obtain reliable ability ratings on the same scale for all participants. Fourth, replication studies can be carried out easily, because new data come in every day, and analyses can thus repeatedly be performed on new data. Fifth, not only participants but also item difficulties can be analyzed, because these are calibrated by the adaptive algorithm as well (Klinkenberg, Straatemeier, & Van der Maas, 2011).

Therefore this method of data collection can be a valuable contribution to the existing literature if analyses are performed in two steps: first, as a method of validation, it is investigated whether effects that are well-known in the literature are also present in these data. If so, the results are valid and analyses can be taken one level further: the number of participants and the amount of data enable more precise analyses than have been carried out in most other studies in the field. Using this approach, we investigated visuospatial working memory, and relations with various mathematical domains in children throughout primary school.

1.1. Visuospatial working memory

Working memory entails the ability to temporarily process and retain a limited amount of information. The degree of separation of different parts of working memory is still debated. An often-used distinction is the tripartite model by Baddeley (Baddeley, 1992; Baddeley & Hitch, 1974). The model contains two slave systems that temporarily store

☆ This research was supported by a grant from the Netherlands Organization for Scientific Research (NWO).
* Corresponding author. Tel.: +31 20 525 6737; fax: +31 20 639 0026.
E-mail addresses: s.h.g.vanderven@uva.nl (S.H.G. van der Ven), h.l.j.vandermaas@uva.nl (H.L.J. van der Maas), m.straatemeier@uva.nl (M. Straatemeier), b.r.j.jansen@uva.nl (B.R.J. Jansen).

1041-6080/$ – see front matter
http://dx.doi.org/10.1016/j.lindif.2013.09.003

verbal and visuospatial information: the phonological loop and the visuospatial sketchpad. The third component is the central executive, responsible for attentional control. The phonological loop and visuospatial sketchpad are commonly measured with simple span tasks, in which a sequence of items (e.g., a dot that appears in different places in a grid) must be replicated. The central executive is measured with complex span tasks: tasks that require remembering a sequence and simultaneously performing another task. For example, the to-be-remembered stimuli must first be counted, or repeated backwards. In factor analyses, the posited tripartite structure has been confirmed in adults and in children (Gathercole, Pickering, Ambridge, & Wearing, 2004), but sometimes the central executive is also divided into a verbal and visuospatial component (e.g., Alloway & Passolunghi, 2010; Friso-Van den Bos, van der Ven, Kroesbergen, & van Luit, 2013).

It has been claimed that backward span tasks, in which a sequence must be reproduced in reversed order, differ substantially from forward span tasks, as only the former tap the central executive. This has been confirmed for verbal tasks (Gathercole et al., 2004), but not for visuospatial tasks. In one study, forward and backward Corsi block (Corsi, 1972) scores loaded on the same factor and there was no difference between forward and backward span size (Kessels, Van den Berg, Ruis, & Brands, 2008). The reason might be executive involvement, also in forward tasks (Fisk & Sharp, 2003; Vandierendonck, Kemps, Fastame, & Szmalec, 2004). Vandierendonck et al. posited that both types of tasks rely on visuospatial storage but also on executive control, which becomes extra involved when the visuospatial store is overloaded. Moreover, Logie (1995) argued for a conceptual division of the visuospatial sketchpad into a visual and a spatial subcomponent. He conceptualized the visual component as a passive store for static, non-moving stimuli, while the spatial component actively rehearses moving, sequential stimuli. The presence of movement requires more executive attentional activation than processing static stimuli does (Rudkin, Pearson, & Logie, 2007).

Together, this suggests that the boundary between tasks designed to measure visuospatial short-term memory and visuospatial tasks tapping the central executive is not as clear-cut as originally posited in the Baddeley model, especially when a task is presented sequentially. The difference between a forward and a backward span is likely small in visuospatial tasks, as both require involvement of the central executive. Therefore in this article we use the term visuospatial working memory to refer to the temporary storage (visuospatial sketchpad) and executive processing of sequential visuospatial information, and apply this term to both backward and forward spatial span tasks.

1.2. Measuring visuospatial working memory

In most visuospatial span tasks, such as the Corsi Block-Tapping Task (Corsi, 1972), or Dot Matrix in the Automated Working Memory Assessment (Alloway, 2007), the number of stimuli in an item gradually increases and a cut-off rule is applied to determine the maximal span length of a participant. Either span length, or the number of items answered correctly, is then used for further analyses or for diagnostic purposes. In this procedure it is implicitly assumed that the number of stimuli in an item is the only factor determining the difficulty of an item. Some studies have already shown, however, that at least for forward sequences other factors also contribute to item difficulty, such as distance between stimuli and position of stimuli in a grid (Bergman Nutley, Söderqvist, Bryde, Humphreys, & Klingberg, 2009; Busch, Farrell, Lisdahl-Medina, & Krikorian, 2005; Parmentier, Elford, & Maybery, 2005), but conclusive results are still lacking, as the focus is usually on few, mostly only one, characteristics at a time.

1.3. Visuospatial working memory and mathematics

There is a growing body of studies showing a strong relationship between working memory abilities and mathematical performance (for a review, see Raghubar, Barnes, & Hecht, 2010). Correlations between visuospatial working memory measures and math performance have almost invariably been found in normally developing children of various ages (e.g., Blair & Razza, 2007; Bull, Espy, & Wiebe, 2008; De Smedt et al., 2009; Swanson & Kim, 2007; Van der Ven, Kroesbergen, Boom, & Leseman, 2012), and children with mathematical difficulties have been found to perform worse on working memory tasks than their normally developing peers (McLean & Hitch, 1999; Van der Sluis, Van der Leij, & De Jong, 2005).

Nevertheless, this relation is not constant. It has occasionally been found that the relationship between math and visuospatial working memory is especially pronounced in young children and decreases as children grow older (De Smedt et al., 2009; Holmes & Adams, 2006; Kyttala, Aunio, & Hautamaki, 2010; Rasmussen & Bisanz, 2005). There are three possible, not mutually exclusive explanations for this finding, which we name development, novelty, and domain specificity.

According to the developmental explanation, young children rely more on visuospatial representations and/or use more visuospatial strategies (Rasmussen & Bisanz, 2005), such as finger counting or mentally counting objects or numbers on a number line. These strategies are later replaced by verbal routines (De Smedt et al., 2009) and retrieval of rote-learned, usually verbal associations between a problem and its answer. Schools usually explicitly teach verbal algorithms, which probably strengthens this developmental effect.

According to the novelty explanation, not so much age is the driving force behind a shift from visuospatial to verbal strategies, but rather novelty of the material. Children of any age and even adults may particularly rely on visuospatial working memory when faced with math problems that are new and challenging for them, while in a later stage they may develop verbal procedures to solve the same problems (Raghubar et al., 2010). When adaptive math tests were administered to 7-year-olds and 8-year-olds, meaning that relative to their ability the test was equally difficult in both age groups, indeed there was no evidence of a decline in importance of visuospatial working memory (Alloway & Passolunghi, 2010).

The domain specificity explanation states that the relation between math and visuospatial working memory differs per math domain. Addition and subtraction may be performed by manipulating quantities, or visualizing these manipulations, while multiplication and its counterpart division, however, are often performed by retrieving facts stored by means of verbal rote memorization (Dehaene & Cohen, 1997; Roussel, Fayol, & Barrouillet, 2002; Zhou et al., 2007). Indeed, brain areas related to visuospatial processing have been shown to be more active when solving simple addition problems, while areas related to verbal processing were more active when solving simple multiplication problems, even in adults, for whom all addition and multiplication problems were very easy (Zhou et al., 2007). As a consequence, relations with visuospatial working memory are expected to be stronger for addition and subtraction than for multiplication and division. Since children start learning addition and subtraction before multiplication and division, domain-specific relations might induce a spurious development effect when visuospatial working memory ability is related to age-appropriate math tests that are not matched for math content.

Thus far, studies have been limited in age range and math domains included. A conclusion regarding the validity of each of the three explanations is therefore still lacking. With Math Garden data are collected in large samples of a wide age range, and scores are obtained for different mathematical domains. This enables a closer analysis of the three different explanations concerning visuospatial working memory and math at different ages.

1.4. Goals of the present study

The goals of the present study are threefold. Our first aim is to validate the Math Garden method of collecting visuospatial working memory data online, without the presence of an experimenter, with computer adaptive testing algorithms. While we acknowledge that this method of collecting data has the disadvantage of less control over the experimental conditions, the possible advantages in terms of sample size are large and therefore this type of data collection potentially forms a valuable contribution to the existing literature, in which smaller data sets are obtained in a more controlled setting. Our approach to validate the data collection method is to test on the item level if the continuously adapted item difficulties of the working memory items are stable over time, and on the child level whether grade differences appear, reflecting the development of working memory during childhood. If we find explainable and stable results, the data collection procedure is sufficiently reliable and valid for further analyses.

The second aim concerns an analysis of the difficulty of spatial working memory items. Math Garden offers a large item bank, which enables a concurrent analysis of different item characteristics. We investigate the relative importance in determining the difficulty of an item of various item characteristics: item length, type of task (forward or backward), grid size, stimulus repetitions (two consecutive stimuli in the same location), stimulus duplicates (two non-consecutive stimuli in the same location), and stimulus location in the grid (in a corner, on the side or elsewhere). In addition, we added control variables that we did not expect to be related to item difficulty: stimulus color and stimulus shape. We expected an item to be more difficult if the average distance between consecutive stimuli is large (Bergman Nutley et al., 2009; Parmentier et al., 2005), if the grid is large, if the sequence must be repeated backwards, if stimuli are presented somewhere in the middle of the grid rather than in a corner or on the side (Bergman Nutley et al., 2009), if there are no repetitions and if there are duplicates. The opposing effects of the final two factors may sound counterintuitive at first, but we expected that while repetitions are relatively easy, duplicates can be confusing: the participant may accidentally skip the stimuli between the two duplicate stimuli.

In order to test whether there is a fundamental difference between forward and backward tasks, we also repeated the regression analyses for all forward and backward items separately, to see if the results are comparable.

The third aim of this study is to investigate the relation between visuospatial working memory and various mathematical domains at different ages throughout primary school. We investigated the size of the relation between visuospatial working memory and the four basic mathematical operations: addition, subtraction, multiplication, and division. If the developmental explanation is true, we expect to see a decline in the relation between working memory and math in higher grades, reflecting a shift from visuospatial to verbal strategies. If the novelty explanation is true, we expect peaks in the relation between working memory and each math domain in the grade in which this domain is introduced: addition and subtraction in grade 1, multiplication in grades 2–3 and division in grades 3–4. If the domain-specificity explanation is true, we expect visuospatial working memory to be especially strongly related to addition and subtraction ability, for which visuospatial strategies are often used, and less to multiplication and division, for which verbal strategies and rote memorization are common strategies.

2. Method

2.1. Participants

The participants of Math Garden are mainly children attending schools that buy accounts for all their pupils. We thus have a sample that is reasonably representative for the Netherlands. We selected the children from the first year of primary education (approximately 6–7 years old) until the 6th and final year of primary education (11–12 years old) that had played the visuospatial working memory task, the Mole Game, within the first year after the game had been introduced.

When children enter Math Garden, a short calibration phase is needed to estimate their ability. For each analysis, therefore, we selected the children that had played at least 30 items of each game involved in that analysis. After this first selection, children were included in each monthly replication of the analyses if they had played at least one item of the analyzed game(s) in that month.

There were 25,954 children who met the criteria to be involved in at least one analysis. Some of these children, however, had been registered by their teachers with highly improbable age–grade combinations (e.g., a fourteen-year-old in grade 1). Since chances are high that for these children mistakes have been made in the registration procedure, we excluded data from these children. We computed the trimmed mean age of each grade (the 10% extremes on each end were trimmed), and excluded data from the children whose reported age deviated more than one year from this trimmed mean age. In addition, we removed data from children with reported dates of birth that were improbably frequent (mainly January 1 of any year). Together, we removed data from 3,225 children (12.4%); the final sample consisted of data from 22,731 children.[1] Note that since individual children did not play the games in every month that we performed the analyses, the sample size differed per analysis. Mean sample sizes for every analysis, as well as age and gender of the participants, are presented in Table 1.

2.2. Materials

2.2.1. Math Garden

2.2.1.1. Math Garden interface. Both the visuospatial working memory game and all math tasks analyzed in this study work according to the same game principles. The games are presented in a garden interface with plants, each representing a game. By playing the games, children's plants grow (the higher the ability, the bigger the plant) and the children earn coins for each solved problem.

If a child clicks on a plant, the corresponding game starts. Children receive an item and are given a limited amount of time, 20 s for most games, to solve it. 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. There is a question mark that the child can click if (s)he does not know the answer: in this case, and also 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 a game earlier or play the game several times.

2.2.1.2. Math Garden computer adaptive technology. By means of a computer adaptive algorithm based on IRT modeling, in Math Garden every instance of a child playing an item is used to adjust the estimates of the ability of the player and the difficulty of item. This is done using a procedure first invented by Elo (1978) and often used in chess. This method is briefly outlined here: for a full description, we refer to Klinkenberg et al. (2011), and for mathematical foundations to Maris and Van der Maas (2012).

Ability of the child and difficulty of the item are both expressed as ratings. The analyses presented in Section 3 are all based on these child ability ratings and item difficulty ratings. The child's expected score on a particular item can be derived from the difference between child ability rating and item difficulty rating, in a way very similar to IRT modeling (Klinkenberg et al., 2011). This expected score depends on the time limit: if the time limit is 20 s, the expected score is

[1] We also performed our analyses without these exclusion criteria, and with a more lenient criterion of a maximum age deviation of 2 years from the trimmed grade mean. This did not change the conclusions.

Table 1
Sample characteristics.

Grade	Total n (boys)	Age[1] M(sd)	Mean n per month[2]			
			+	−	x	/
1	3,681 (2,046)	7.2 (0.38)	988.7	765.2	406.3	155.4
2	4,868 (2,531)	8.2 (0.40)	1266.6	1209.4	903.2	329.7
3	4,916 (2,655)	9.3 (0.42)	1087.7	1043.7	994.2	587.4
4	4,041 (2,171)	10.3 (0.42)	805.2	754.8	778.2	619.7
5	3,104 (1,737)	11.3 (0.43)	537.2	501.0	528.5	479.2
6	2,121 (1,175)	12.3 (0.43)	385.3	356.6	372.3	360.8
Total	22,731 (15,956)	9.5 (1.64)				

[1] Age at the end of the data collection.
[2] Children that played both the working memory game and the respective math game in that month.

somewhere between −1 (immediate wrong answer) and 1 (immediate correct answer). After the child played the item, the expected score and the obtained score are compared. The child's ability rating is then adjusted upward if the child scored higher than expected: apparently the child's ability is better than previously thought. The ability is adjusted downward if the child scored lower than expected. The larger the discrepancy between the true and expected score, the more the child's rating is adjusted. In a similar fashion, the item difficulty rating is also adjusted: downward if the child scored higher than expected (apparently the item was easier than previously thought) and upward if the child scored lower than expected.

New children and new items enter the system with an estimated start rating. Because of the updating procedure, these ratings are adjusted after every trial, and they tend to approach their true value quite quickly (Klinkenberg et al., 2011).

While in the original Elo (1978) method only wins and losses are considered, and, similarly, standard IRT models are based on accuracy only, the Math Garden algorithms are novel in also taking response time into account. Math Garden uses a scoring rule for response times that has strong psychometric properties, solves the speed–accuracy trade-off problem, and with its visualization of won or lost coins it has proven easy to understand for children (Klinkenberg et al., 2011; Maris & Van der Maas, 2012).

The item selection procedure is also based on item difficulty rating and child ability rating. Based on the child ability rating, the item difficulty rating is determined at which the child has a probability of .75 of finding the correct answer within the time limit. The next item is sampled around this difficulty level ($M = .75$,[2] $SD = .10$). In order to prevent fast repetition (and thereby recognition) of the same items, the last 20 administered items are excluded in the item selection procedure. Children can also choose to solve easy problems, with an average probability of .90, or difficult problems, in which this probability is .60.

2.2.2. Visuospatial working memory: Mole Game

In a grid with molehills, varying in size from 3 × 3 to 5 × 5, moles appear sequentially. Each mole carries a banner with a colored (yellow, blue, red or green) shape (circle, square or triangle). After the sequence, the child is told to indicate the sequence of the moles, in the same or reversed order. An example item is shown in Fig. 1. The item bank of the Mole Game consists of 360 items, ranging in length from 1 to 9 moles, with 40 items of each length. Half of the items of each length are forward items, in which the mole locations have to be repeated in the same order. The other half of the items are backward items, in which the mole locations have to be repeated in reverse order. As sequences with only one mole cannot be classified as forward or backward, and some of the other item characteristics outlined below do not apply to items with one mole either, we did not analyze those items, meaning that there were 320 items left.

We looked at the following item characteristics of these 320 items:

- Sequence length (number of stimuli)
- Grid size: the number of squares where moles could potentially appear. The grid could be 3 × 3, 3 × 4, 4 × 4, 4 × 5, or 5 × 5 molehills.
- Mean inter-stimulus distance, measured in number of squares (Euclidean distance[3]).
- Proportion of different colors on the banners: (Number of different colors −1)/(stimulus length −1). The 1 was subtracted to correct for the first item, which always has a color, but is not *different* yet.
- Proportion of different shapes on the banners: (Number of different shapes −1)/(stimulus length −1)
- Proportion of repetitions (two consecutive stimuli in the same location): number of repetitions/(number of stimuli −1)
- Proportion of duplicates (two non-consecutive stimuli in the same location): number of duplicates/(number of stimuli −1)
- Proportion of stimuli appearing in a corner
- Proportion of stimuli appearing on the side
- Proportion of stimuli appearing exactly in the center
- Task (forward or backward)

2.2.3. Math games

In this paper, we used the data from four math games: the Addition, Subtraction, Multiplication, and Division game. All games work according to the Math Garden principles outlined above. In the Addition, Subtraction, Multiplication and Division games, problems with the respective basic operators are presented. The Addition and Subtraction games are answered in a multiple choice format; the Multiplication and Division games present open questions. To prevent young children from being faced with operations that they are unable to perform, the Multiplication and Division games are only present in the garden interface if the child's ability on the Addition and Subtraction games is sufficiently high. The Math Garden math games have good criterion validity (Klinkenberg et al., 2011): they correlate strongly with the Dutch norm-referenced math tests created by Cito (Janssen, Scheltens, & Kraemer, 2005) that are administered twice a year in almost every Dutch primary school.

3. Results

3.1. Validation of the Mole Game

To validate the reliability and validity of the Mole Game, we performed two analyses: we tested the stability of item difficulty ratings, and grade differences in children's ability rating. Moreover, whenever possible, also while addressing the second and third aims, we performed the same analyses repeatedly, to see if all results were stable. Repetitions of the same analyses were performed weekly for item difficulty rating analyses and monthly for child ability analyses. These different repetition windows were chosen because individual children played less frequently than each item was being played (since $N_{players} >> N_{items}$).

3.1.1. Item difficulty rating stability

We obtained item difficulty ratings of each item of the Mole Game on a weekly basis, starting on the day the task was introduced in Math Garden, and ending 46 weeks later. Thus, we have 46 difficulty ratings of every item. Since the Elo updating system (Elo, 1978) involves an adjustment of the old rating, initially item ratings were still influenced by

[2] Normally in adaptive testing, a probability of .50 is maximally informative. This is, however, experienced as discouragingly low by test takers. Since response latency is included in the scoring method as well, it is possible to present easier problems, yet obtain sufficient and reliable information to determine players' abilities (see section 3.1 of Klinkenberg et al., 2011, for a more thorough discussion).

[3] We also performed the analyses with Chebyshev distance (the number of horizontal, vertical and/or diagonal spaces) and City Block distance (the number of horizontal and vertical spaces). The different distance measures were highly correlated, and this did not influence the results.

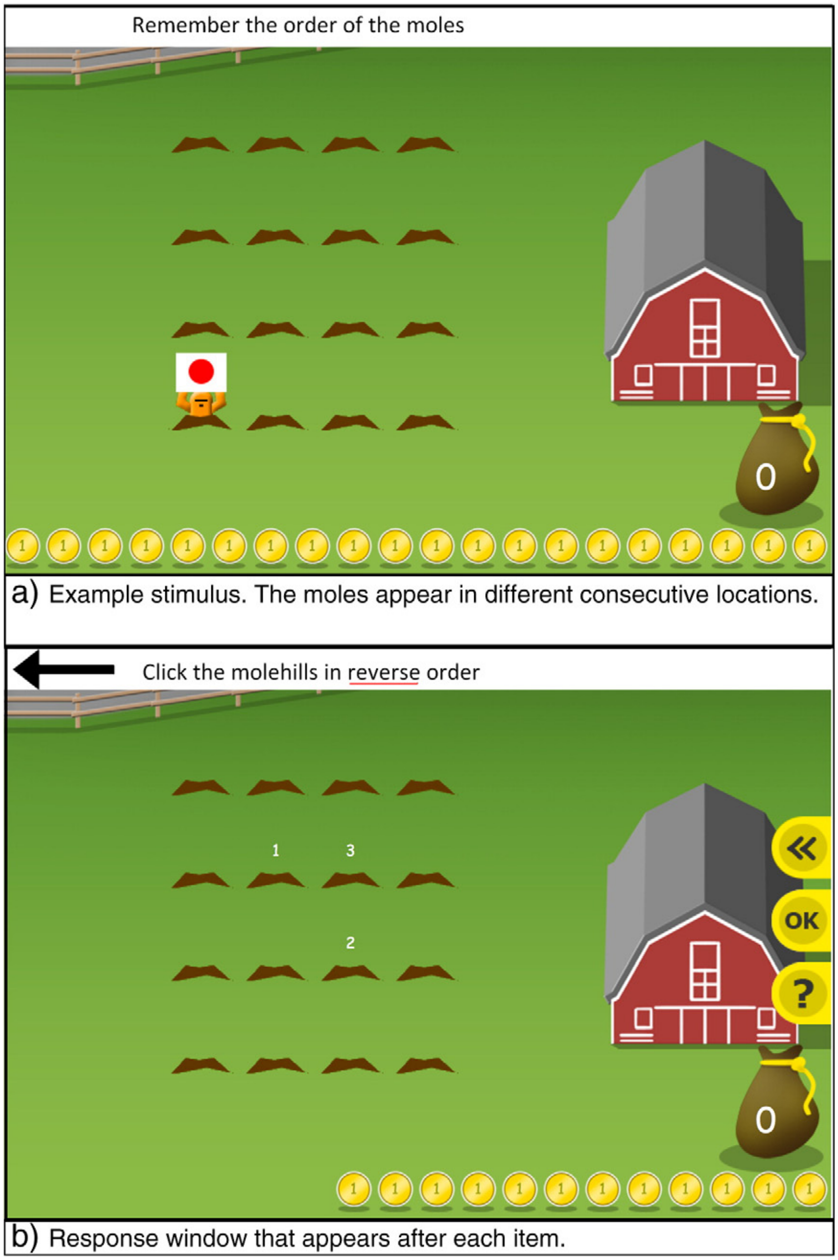

Fig. 1. Mole Game.

their starting value, which was determined by us and mainly based on the number of stimuli and type of task (forward or backward). Some items, however, proved to be so difficult that they were hardly ever or even never played. The difficulty ratings of these items are still very close or equal to their relatively arbitrary start ratings, and these items should therefore be excluded from the item rating analyses. Therefore we only selected the items that had been played at least 250 times[4] within the 46 weeks of the analyses. This procedure ensured that in the final analyses, the item difficulty ratings could no longer be affected by the rather arbitrary start ratings. Of the 320 items in the bank, 194 items (106 forward items, 88 backward items) met this criterion (range: 268 to 23,608 times played, $M = 15{,}042$). The items that had not been played sufficiently frequently were too difficult: none of the forward items with eight or more moles, and none of the backward items with seven or more moles, had been played sufficiently often. Even for the most skilled players there was no ceiling effect in the task: 62 items had such a high start rating that they had not even been presented once.

Fig. 2 shows the development of the item difficulty ratings over the course of the 46 weeks, for 39 of the 194 items. Including more items would impair the visibility of the individual trajectories in Fig. 2; the 39 items are a representative selection of the different trajectories. The Figure shows how the item difficulty ratings move from their initial, arbitrarily chosen value to their empirically determined value. Some lines are not yet present at time zero, because items are included from

[4] We repeated the analyses with other boundary criteria: minima of 25 and 500. The results were very similar.

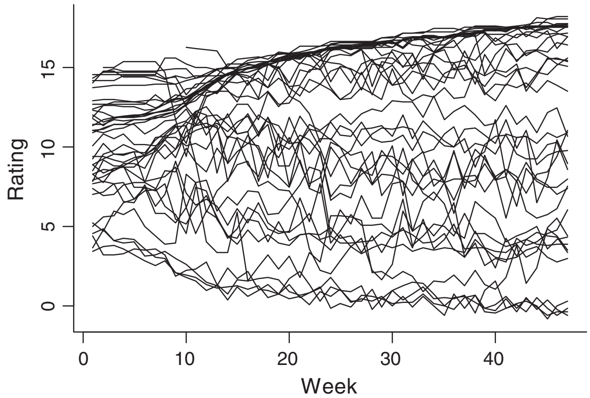

Fig. 2. Development of item ratings over time.

the week they had been played for the first time. From week 15 onwards, all 194 items had been played at least once.

Stability of the item difficulty ratings was investigated in two ways. First, for each individual item, the item auto-correlation was determined with a lag of one week. This analysis showed that the items are reasonably stable in their difficulty ratings: the mean individual item auto-correlation with a lag of one week was .79 (SD = 0.21). Second, inter-item correlations were determined between all item difficulty ratings together, with ten-week intervals. These were very high, with *r*-values between .92 and .99, all *ps* < .001.

3.1.2. Grade differences in visuospatial working memory

As a second step of validation of the Mole Game, we investigated visuospatial working memory ratings in different grades. These analyses were performed monthly. The mean rating of the Mole Game in each grade is shown in Fig. 3. Every month, a one-way ANOVA showed a strong, significant effect of grade: older children obtained significantly higher ratings of visuospatial memory. Test results varied between $F(5,1179) = 82.4$, $p < .001$ and $F(5,8022) = 652.3$, $p < .001$. Post-hoc Tukey HSD tests showed that differences between consecutive grades were always significant, except for the difference between grades 4 and 5, which was significant in only three of the twelve months, and the difference between grades 5 and 6, which was significant in nine of the twelve months. This is consistent with a pattern of decelerated growth over the years, as can also be seen in Fig. 3: differences between consecutive grades diminished in the higher grades.

The results of the item analyses and grade differences showed a stable and predictable pattern of results, thus validating the use of the Mole Game. Therefore the second and third aims could be addressed.

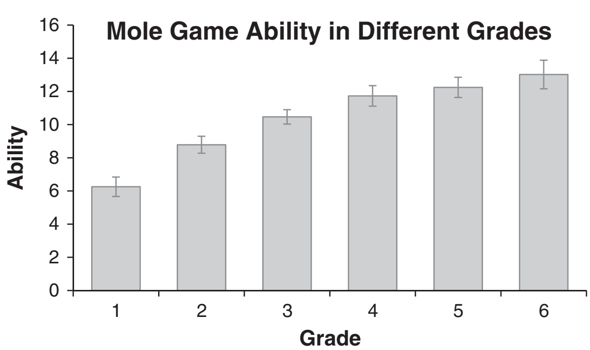

Fig. 3. Mean working memory ratings in every grade, determined every month. Error bars represent the standard deviations of these monthly determined means.

3.2. Item difficulty predicted by item characteristics

In order to address the second aim of the paper, the relative contributions of different item characteristics to item difficulty, three regression analyses were performed: forward and backward items were analyzed separately and together, including type of task as one of the characteristics in the latter analysis. The results of the regression analyses performed in the final week are presented in Table 2. All effects with a t-value outside the two boundary values of −1.97 and 1.97 were significant with an alpha of .05. In all three analyses, item length was by far the strongest predictor. In the forward items, there was a negative effect of duplicates and a positive effect of inter-stimulus distance on item difficulty: the presence of duplicates made an item easier, while a large inter-stimulus distance led to a more difficult item. All other variables, including the variables related to the positions of the stimuli in the grid and the control variables, were not significant. The result of the backward items analysis was similar to the forward items, with some exceptions: there was also a significant effect of repetitions, which made an item significantly easier, a significant effect of grid size (marginally significant in the forward condition) but the duplicate effect was reversed: duplicates made an item more difficult in the backwards condition. This reversal effect is remarkable. In the combined analysis of forward and backward items, therefore, two extra predictors were added: the type of task, and an interaction effect of task × duplicates. Again, item length was the strongest predictor of item difficulty. The presence of duplicates and repetitions led to easier items, while a larger inter-stimulus distance and a larger grid size led to more difficult items. Backward items were more difficult than forward items and the task × duplicates interaction was also significant: the presence of duplicates in backward items led to a higher item difficulty. The other variables, again including the position of the items in the grid and the control variables were not significant.

To test stability of these results, the same analysis was repeated every week, for 46 weeks. The development of the t-values over time of the combined analysis is depicted in Fig. 4. After the first few weeks, in which item ratings moved from the starting values towards their true ratings, one would expect the effects of each predictor to stabilize over time (i.e., lines in the Figure becoming horizontal). This is indeed what Fig. 4 shows: after approximately 10 weeks, the effects of the predictors stabilize.

3.3. Visuospatial working memory and math: domain and grade differences

As a third aim, we looked at the relation between visuospatial working memory performance and math ability in different math domains, and we investigated age differences in this relation. Within each grade, regression analyses were performed to investigate whether the Mole Game predicted math ability in the different math domains. The analyses were again repeated monthly to test the stability of the results. Therefore a total of 6 grades * 12 months * 4 math domains = 288 analyses were performed. However, when an analysis consisted of fewer than 50 children, this analysis was removed, leading to a total of 281 analyses. The explained variance in the regression analyses is depicted in Fig. 5. The results show a consistently significant relation between visuospatial working memory and the various math domains. Of the total of 281 investigated relations, 275 were significant, the largest exceptions being tasks that children had not learned yet in school (e.g., multiplication in grade 1, division in grades 1 and 2). Fig. 5 also shows that the size of the relation between visuospatial working memory and math differs across the math domains: addition and subtraction show the strongest relations while the relation with multiplication and division is smaller.

Fig. 5 also suggests that the relationship between visuospatial working memory and math differs by grade. Especially the addition graph suggests that the relation is largest in the higher grades, and then diminishes in later grades. With moderation analyses (Baron & Kenny, 1986)

Table 2
Results of multiple regression analyses in the final week.

	Forward items only (n = 106)				Backward items only (n = 88)				All items (n = 194)			
	R^2 = .88				R^2 = .90				R^2 = .86			
	β	SE	t	p	β	SE	t	p	β	SE	t	p
Task	–	–	–	–	–	–	–	–	**.11**	**0.03**	**3.44**	**<.001**
Length	**.94**	**0.05**	**19.45**	**<.001**	**.86**	**0.04**	**19.96**	**<.001**	**.90**	**0.03**	**28.41**	**<.001**
Grid size	*.11*	*0.07*	*1.67*	*.098*	**.15**	**0.05**	**3.33**	**.001**	**.14**	**0.04**	**3.51**	**<.001**
Distance	**.15**	**0.06**	**2.49**	**.015**	**.17**	**0.05**	**3.18**	**.002**	**.16**	**0.04**	**3.82**	**<.001**
Colors	.06	0.05	1.29	.201	.02	0.04	0.55	.581	.05	0.03	1.55	.123
Shapes	−.02	0.04	−0.51	.612	.04	0.04	1.05	.297	.00	0.03	−0.01	.994
Repetitions	−.04	0.05	−0.84	.404	**−.13**	**0.05**	**−2.74**	**.008**	**−.08**	**0.34**	**−2.43**	**.016**
Duplicates	**−.12**	**0.05**	**−2.40**	**.018**	**.14**	**0.04**	**3.24**	**.002**	**−.11**	**0.04**	**−2.62**	**<.001**
Corners	−.06	0.07	−0.76	.451	−.01	0.05	−0.28	.780	−.03	0.04	−0.67	.502
Side	.01	0.06	0.09	.931	.04	0.05	0.88	.379	.03	0.04	0.79	.430
Center	.01	0.04	0.19	.847	−.03	0.04	−.70	.488	−.01	0.03	−0.33	.741
Duplicates*task	–	–	–	–	–	–	–	–	**.16**	**0.03**	**3.98**	**<.001**

bold = significant (p < .05).

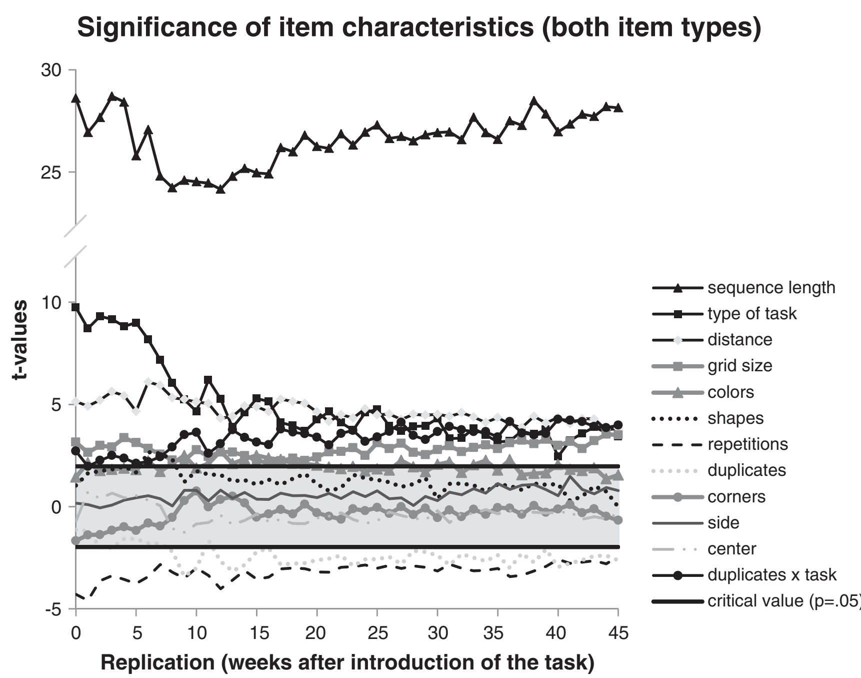

Fig. 4. Results of regression analyses over time: weekly replications of the same analysis. Values outside the light gray area are significant; the further away from the gray, the stronger the effect of the corresponding predictor.

with grade as a moderator, it was tested whether the relation between visuospatial working memory and math changed significantly in higher grades. First, visuospatial working memory ratings and math ratings were standardized in each grade separately.[5] Then a series of regression analyses was performed, in which the ratings in each math domain were regressed on visuospatial working memory and the interaction effect of visuospatial working memory × grade. The main effect of grade was not included because visuospatial working memory and math ratings were standardized by grade. A significant interaction effect, while the other variables are standardized, means that the strength of the main effect changes across grades. The results of these analyses in the final month are presented in Table 3. Grade was coded such that grade 1 was 0, and each subsequent grade was coded 1 higher. As such, the main effect of visuospatial working memory represents the estimated strength of relationship between visuospatial working memory and math in grade 1. The interaction effect shows the change in this relationship if grade increases by 1. For example, the relationship between working memory and addition is .54, so a 1 *SD* increase in working memory rating is accompanied by a .54 *SD* increase in addition rating in grade 1. The interaction effect is estimated at −.03, which means that the size of this relationship decreases by .03 in each subsequent grade. For example, a 1 *SD* increase in VWM rating in grade 5 is accompanied by only a .54 – 5 *.03 = .39 *SD* increase in addition rating. A decreasing effect of working memory on math was observed for the Addition and Subtraction games. For the Multiplication game there was no significant interaction effect, and for the Division game the interaction effect was significant in the opposite direction: an increasingly stronger relation with visuospatial working memory in the higher grades. However, Table 1 shows that relatively few children played the Multiplication game in grade 1 and the Division game in grades 1 and 2 (and grade 3 to a lesser degree). Since these games become available only with a minimum addition and subtraction ability, there is likely a selection bias in these data. Three repetitions of the analyses

[5] Standardization was done for each grade separately to correct for a potential age confound: children in higher grades have higher ratings in both working memory and mathematics.

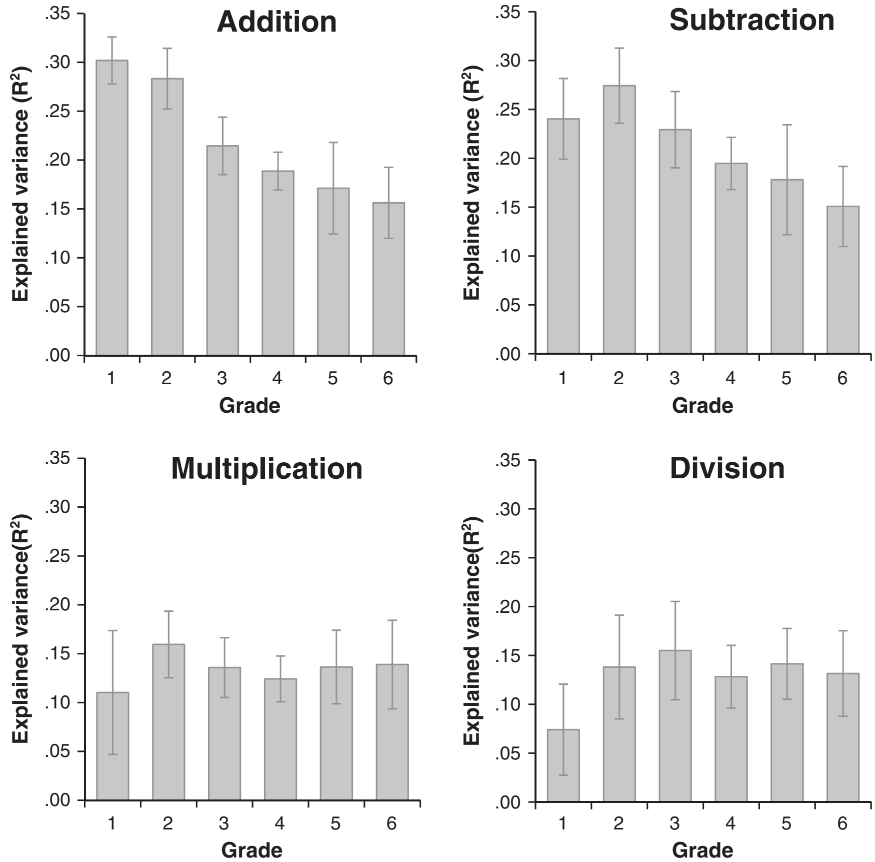

Fig. 5. Explained variance in mathematics by the Mole Game. Error bars represent the standard deviations of these monthly determined explained variances.

(excluding grade 1, grades 1–2 and grades 1–3 respectively) did not change the results for multiplication, but in all three repetitions there was no longer a significant visuospatial working memory × grade interaction effect for division either.

These analyses were also performed monthly. The rightmost column of Table 3 shows the number of months in which the results of the analyses were the same as those presented fully in the Table.

4. Discussion

With the present study, we had three aims. The first aim was establishing the validity of the method of measuring visuospatial working memory ability with the Math Garden interface and algorithms, in which the ability of the participants and the difficulty of the items is estimated on the fly, using a computer adaptive testing procedure based on the Elo rating system. While the mathematical tasks using the same algorithm had already been validated in a previous study (Klinkenberg et al., 2011), the visuospatial working memory game, known as the Mole Game, had been newly added. The second aim was to analyze how different item characteristics of a visuospatial working memory task contribute to item difficulty. The third aim was to investigate the interrelations between working memory and various mathematical domains at different ages, and see if differences can be explained by development, novelty, and/or the domain specificity.

Regarding the first aim, the item difficulty ratings were stable over time and there was a significant age effect in visuospatial working memory ability: older children had higher ratings. This is consistent with previous literature showing age-related development in working memory (e.g., Alloway, Gathercole, & Pickering, 2006; Huizinga, Dolan, & Van der Molen, 2006; Van der Ven et al., 2012). This confirmed that the Mole Game is a suitable game and that the Math Garden algorithms can also be used for measuring visuospatial working memory. The results therefore show that Math Garden is a valuable tool to study children's development, not only in mathematics but also in other, related cognitive domains. The tool enables a thorough item analysis of visuospatial working memory, and reliably measures performance in different mathematical domains, in a far larger age and ability range than commonly employed in research designs. However, the data were collected online, which means that the degree of control over playing frequency and playing circumstances is less than in traditional research. On the other hand, test circumstances are ecologically more valid, as one measures math while children learn and practice in the classroom, rather than in the artificial environment of the lab setting.

The on the fly item calibration algorithms enabled an analysis of the visuospatial working memory item difficulties: the second aim of this paper. The analyses showed that item difficulty ratings could be predicted well from the item characteristics expected to play a role in item difficulty, while control variables, not expected to be related to item difficulty, were indeed not significant. Item length was by far the best predictor of item difficulty, and backward items were more difficult than forward items. Nevertheless, some other characteristics were also significantly related to item difficulty: the size of the grid in which the stimuli appeared, the interstimulus distance, and the presence of repetitions or duplicates (stimuli appearing in the same location). Grid size may reflect the visuospatial counterpart of a process described in verbal working memory: redintegration (Brown & Hulme, 1995; Hulme et al., 1997). This process entails that a memory trace of a word that slowly decays, can be restored to full strength with phonological and semantic knowledge. In visuospatial working memory, if knowledge of a stimulus location has decayed partially, i.e., if a participant only knows the approximate location of the stimulus, in a small grid the exact location

Table 3
Results of the moderation analysis: visuospatial working memory and math in different grades.

	B	SE B	R^2	ΔR^2	Similar results:
Addition					
Step 1			.244		11 out of 12 analyses
VWM	.49**	.01			
Step 2			.246	.002***	
VWM	.54***	.02			
VWM x grade	−.03***	.008			
Subtraction					
Step 1			.260		8 out of 12 analyses
VWM	.51***	.01			
Step 2			.261	.001***	
VWM	.56***	.02			
VWM x grade	−.03***	.01			
Multiplication					
Step 1			.173		9 out of 12 analyses
VWM	.41***	.01			
Step 2			.173	.000	
VWM	.40***	.02			
VWM x grade	.003	.001			
Division I (all grades)					
Step 1			.166		6 out of 12 analyses
VWM	.42***				
Step 2			.169	.003**	
VWM	.34***				
VWM x grade	.03**				
Division II (excluding grade 1 and 2)					
Step 1					11 out of 12 analyses
VWM	.44***		.179		
Step 2					
VWM	.47***		.179	<.001	
VWM x grade	−.01				

* $p < .05$, ** $p < .01$, *** $p < .001$ VWM = visuospatial working memory (Mole Game).

can be reconstructed, but in a larger grid there are more competing locations close to this approximate location. The effect of interstimulus distance may reflect a chunking process. Although the exact number of chunks is debated, it is generally agreed upon that combining multiple stimuli into a single chunk increases the capacity of working memory (Cowan, 2001). Stimuli that are close together in space may be more easily coded as a single chunk.

The results also showed that backward tasks were significantly more difficult than forward tasks, but this difference was not large and the pattern of predictors of item difficulty was comparable for forward and backward items. This further confirms that there is no clear-cut distinction between forward and backward spatial span tasks, possibly because also forward visuospatial tasks require active executive involvement (Fisk & Sharp, 2003; Logie, 1995; Rudkin et al., 2007; Vandierendonck et al., 2004). An alternative explanation is that for verbal tasks, a backward task qualitatively changes the processes that are involved; there are indications that backward recall of a verbal span task requires additional visuospatial processing compared to forward recall, especially in those who report using visual imagery to reverse the words (St Clair-Thompson & Allen, 2013). Coordinating these processes may require additional executive involvement for a backward task. Given that a visuospatial task is already visuospatial by itself, qualitative differences between forward and backward visuospatial span tasks may be less pronounced. From an ecological point of view the lack of large differences in a visuospatial task is also understandable; in real life there is no need to reverse verbal phrases, which is therefore effortful, while reversing visuospatial information is far more natural, e.g., following a route in a backwards direction, or retracing a moving object to find its origin. This suggests that contrary to verbal working memory, for visuospatial working memory a backwards span task may not be the most suitable task to tap the central executive.

Despite the similarities, there were also some differences between the importance of item characteristics in forward and backward tasks, notably the presence of duplicates (two non-consecutive stimuli appearing in the same location). The presence of duplicates made forward items significantly easier, but backward items significantly more difficult, and in the overall analysis there was a significant interaction effect of task × duplicates. The reason may be that especially for backwards items, it is easier to be led astray: if one also has to reverse the sequence it might be more tempting to skip the items between the two duplicate items.

Together, this knowledge can be helpful to construct better working memory tests, leading to better normal distributions of working memory ability in future studies. Often, the distribution is rather 'stepwise', because a number of items of the same length are administered, and then one stimulus is added. This leads to a relatively large number of participants failing from this longer item onwards. Changing item difficulties more subtly than adding a new stimulus, by changing other characteristics of the item, such as interstimulus distance and grid size, may lead to a smoother distribution. Moreover, in young children the variance in visuospatial working memory is often low when traditional tests are used, because adding an extra stimulus to the item length means a large increase in difficulty, especially when a backward test is used. The Math Garden algorithm acknowledges fine-grained differences in item difficulty ratings, which enables a more precise estimation of the children's abilities.

Concerning the third aim, investigating visuospatial working memory and different math domains at different ages, we found significant relations between visuospatial working memory abilities and math abilities, also consistent with previous literature (De Smedt et al., 2009; Lee et al., 2012; Swanson & Kim, 2007). We found a decreasing age trend in addition and subtraction. In higher grades, working memory explained less variance in math performance in these games. This trend was absent for the Multiplication game and Division game (when excluding grade 1). These results support the development and domain-specificity explanations: the developmental explanation seems valid, but restricted to specific domains. In addition and subtraction, young children intuitively develop, and are also taught explicitly, procedural strategies involving quantity manipulation. These strategies pose a high load on visuospatial working memory, but they are later replaced by verbal strategies and retrieval (De Smedt et al., 2009; Rasmussen & Bisanz, 2005; Roussel et al., 2002). Note that mere rote memorization of these problems cannot explain these findings. The problems in Math Garden are presented adaptively, so although a certain math domain becomes more familiar over time, the specific problems that children solve are of equal difficulty to every child, regardless of ability, and problems that children have memorized are too easy to be selected with the adaptive algorithm.

We found no evidence for an age trend in multiplication and division: relations with visuospatial working memory were significant but with lower effect sizes and the size of the relationship did not change in higher grades. This may reflect that in multiplication education, already from the beginning more emphasis is placed on verbal and retrieval strategies: verbal rote memorization is applied in many schools to learn the multiplication facts (Roussel et al., 2002; Zhou et al., 2007). For division an opposite age trend was found in the first analyses: the relation between working memory and the Division game was larger in higher grades. Subsequent analyses showed, however, that this effect disappeared when the low grades were excluded. Children can only play the Division game if their addition, subtraction and multiplication ratings are sufficiently high. A comparison of the sample sizes, as shown in Table 1, shows that the number of children playing the Multiplication and Division games, compared to the number of children playing the Addition game, is clearly lower for multiplication in grade 1 and for division in grades 1 and 2: the grades when these operations have not been introduced in the curriculum yet. This shows that on average the introduction of the games approximately matches the curriculum of the children, while leaving room for mathematically precocious children to start earlier. The multiplication and division data from these early years should be interpreted with caution because of this selection bias. It must be noted, though, that the strongest decline in strength of relationship for addition and subtraction takes place precisely during these early years. This may partially explain the domain-specific effects.

There is also some evidence supporting the novelty explanation, which predicts peaks in the relation between visuospatial working memory and each math domain in the grade in which the domain is introduced: for the Addition and Subtraction Games in grade 1, for the Multiplication game in grade 2–3 and for the Division game in grade 3–4. Fig. 5 shows slight peaks around the predicted grade: for the Addition Game in grade 1, for the Subtraction Game and the Multiplication Game in grade 2 and for the Division Game in grade 3. Note, however, that especially the Multiplication and Division peaks are low, and even at peak level relations between visuospatial working memory and these domains are far smaller than relations with addition and subtraction.

Together, these results show that the load on visuospatial working memory may be especially high for addition and subtraction problems at younger ages. These differential findings for the different mathematical domains and ages suggest that low visuospatial working memory abilities are a risk factor for mathematical development, especially in the earlier years of primary education. This is so for two reasons: first, because in the earlier years the math curriculum mainly consists of addition and subtraction, which showed the largest relation with visuospatial working memory, and second, in the earlier years this relation is strongest. Children with low visuospatial working memory abilities are likely to struggle with visualizing quantities in early grades, but apparently this can be overcome in later years, as the diminishing relation with visuospatial working memory shows. It must be noted, though, that mathematics comprises a wide variety of skills and the Math Garden games that are analyzed in the present paper represent only a part of this: mental arithmetic ability in the four main operations.

To summarize, this paper showed that it is possible to collect large amounts of data with an online computer adaptive practice program. With this method of data collection it was possible to show that there are more similarities than differences between forward and backward visuospatial working memory items, and that visuospatial working memory was related to mathematical performance, but this relationship was not the same in each math domain. Especially in the domains addition and subtraction, relations with visuospatial working memory appeared, but this relationship diminished in strength in the higher grades.

References

Alloway, T. P. (2007). *Automated Working Memory Assessment.* London: Pearson Assessment (Translated and reproduced by permission of Pearson Assessment).

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. http://dx.doi.org/10.1111/j.1467-8624.2006.00968.x.

Alloway, T. P., & Passolunghi, M. C. (2010). The relationship between working memory, IQ, and mathematical skills in children. *Learning and Individual Differences*, *21*, 133–137. http://dx.doi.org/10.1016/j.lindif.2010.09.013.

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: Academic Press.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182. http://dx.doi.org/10.1037/0022-3514.51.6.1173.

Bergman Nutley, S., Söderqvist, S., Bryde, S., Humphreys, K., & Klingberg, T. (2009). Measuring working memory capacity with greater precision in the lower capacity ranges. *Developmental Neuropsychology*, *35*, 81–95. http://dx.doi.org/10.1080/87565640903325741.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*, 647–663. http://dx.doi.org/10.1111/j.1467-8624.2007.01019.x.

Brown, G. D. A., & Hulme, C. (1995). Modeling item length effects in memory span — no rehearsal needed. *Journal of Memory and Language*, *34*, 594–621. http://dx.doi.org/10.1006/jmla.1995.1027.

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology*, *33*, 205–228. http://dx.doi.org/10.1080/87565640801982312.

Busch, R. M., Farrell, K., Lisdahl-Medina, K., & Krikorian, R. (2005). Corsi block-tapping task performance as a function of path configuration. *Journal of Clinical and Experimental Neuropsychology*, *27*, 127–134. http://dx.doi.org/10.1080/138033990513681.

Corsi, P. M. (1972). *Human memory and the medial temporal regions of the brain.* Unpublished doctoral dissertation, McGill University, Montreal, Canada.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*, 87–185. http://dx.doi.org/10.1017/s0140525x01003922.

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. http://dx.doi.org/10.1016/j.jecp.2009.01.004.

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex*, *33*, 219–250. http://dx.doi.org/10.1016/S0010-9452(08)70002-9.

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

Fisk, J. E., & Sharp, C. A. (2003). The role of the executive system in visuo-spatial memory functioning. *Brain and Cognition*, *52*, 364–381. http://dx.doi.org/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*. http://dx.doi.org/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. http://dx.doi.org/10.1037/0012-1649.40.2.177.

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. http://dx.doi.org/10.1080/01443410500341056.

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*, F9–F15. http://dx.doi.org/10.1111/j.1467-7687.2009.00848.x.

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. http://dx.doi.org/10.1016/j.neuropsychologia.2006.01.010.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1217–1232. http://dx.doi.org/10.1037/0278-7393.23.5.1217.

Jansen, B. R. J., Louwerse, J., Straatemeier, M., Van der Ven, S. H. G., Klinkenberg, S., & Van der Maas, H. L. J. (2013). The influence of practicing maths with a computer-adaptive program on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences*, *24*, 190–197.

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. http://dx.doi.org/10.1177/1073191108315611.

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlstrom, K., et al. (2005). Computerized training of working memory in children with ADHD — A randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, *44*, 177–186. http://dx.doi.org/10.1097/00004583-200502000-00010.

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 in Education*, *57*, 1813–1824. http://dx.doi.org/10.1016/j.compedu.2011.02.003.

Kyttala, M., Aunio, P., & Hautamaki, J. (2010). Working memory resources in young children with mathematical difficulties. *Scandinavian Journal of Psychology*, *51*, 1–15. http://dx.doi.org/10.1111/j.1467-9450.2009.00736.x.

Lee, K., Ng, S. F., Pe, M. L., Ang, S. Y., Hasshim, M. N. A.M., & Bull, R. (2012). The cognitive underpinnings of emerging mathematical skills: Executive functioning, patterns, numeracy, and arithmetic. *British Journal of Educational Psychology*, *82*, 82–99. http://dx.doi.org/10.1111/j.2044-8279.2010.02016.x.

Logie, R. H. (1995). *Visuo-spatial working memory.* Hove, United Kingdom: Lawrence Erlbaum Associates.

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

McLean, J. F., & Hitch, G. J. (1999). Working memory impairments in children with specific arithmetic learning difficulties. *Journal of Experimental Child Psychology*, *74*, 240–260. http://dx.doi.org/10.1006/jecp.1999.2516.

Parmentier, F. B. R., Elford, G., & Maybery, M. (2005). Transitional information in spatial serial memory: Path characteristics affect recall performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 412–427. http://dx.doi.org/10.1037/0278-7393.31.3.412.

Raghubar, K. P., Barnes, M.A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, *20*, 110–122. http://dx.doi.org/10.1016/j.lindif.2009.10.005.

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

Roussel, J. -L., Fayol, M., & Barrouillet, P. (2002). Procedural vs. direct retrieval strategies in arithmetic: A comparison between additive and multiplicative problem solving. *European Journal of Cognitive Psychology*, *14*, 61–104. http://dx.doi.org/10.1080/09541440042000115.

Rudkin, S. J., Pearson, D.G., & Logie, R. H. (2007). Executive processes in visual and spatial working memory tasks. *Quarterly Journal of Experimental Psychology*, *60*, 79–100. http://dx.doi.org/10.1080/17470210600587976.

St Clair-Thompson, H. L., & Allen, R. J. (2013). Are forward and backward recall the same? A dual-task study of digit recall. *Memory & Cognition*, *41*, 519–532. http://dx.doi.org/10.3758/s13421-012-0277-2.

Swanson, L., & Kim, K. (2007). Working memory, short-term memory, and naming speed as predictors of children's mathematical performance. *Intelligence*, *35*, 151–168. http://dx.doi.org/10.1016/j.intell.2006.07.001.

Van der Sluis, S., Van der Leij, A., & De Jong, P. F. (2005). Working memory in Dutch children with reading and arithmetic related learning LD. *Journal of Learning Disabilities*, *38*, 207–221. http://dx.doi.org/10.1177/00222194050380030301.

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. http://dx.doi.org/10.1111/j.2044-8279.2011.02035.x.

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. http://dx.doi.org/10.1348/000712604322779460.

Zhou, X., Chen, C., Zang, Y., Dong, Q., Chen, C., Qiao, S., et al. (2007). Dissociated brain organization for single-digit addition and multiplication. *NeuroImage*, *35*, 871–880. http://dx.doi.org/10.1016/j.neuroimage.2006.12.017.