Psychological Research
DOI 10.1007/s00426-014-0614-0

ORIGINAL ARTICLE

Effect of working memory training on working memory, arithmetic and following instructions

Sissela Bergman-Nutley · Torkel Klingberg

Received: 28 May 2013 / Accepted: 18 September 2014

Abstract Mathematical ability is dependent on specific mathematical training but also associated with a range of cognitive factors, including working memory (WM) capacity. Previous studies have shown that WM training leads to improvement in non-trained WM tasks, but the results regarding transfer to mathematics are inconclusive. In the present study, 176 children with WM deficits, aged 7–15 years performed 5 weeks of WM training. During the training period, they were assessed five times with a test of complex WM (the Odd One Out), a test of remembering and following instructions and a test of arithmetic. The improvements were compared to the performance of a control group of 304 typically developing children aged 7–15 years who performed the same transfer tasks at the same time intervals, but without training. The training group improved significantly more than the control group on all three transfer tests (all $p < 0.0001$), after correction for baseline performance, age and sex. The effect size for mathematics was small and the effect sizes for the WM tasks were moderate to large. The transfer increased linearly with the amount of training time and correlated with the amount of improvement on the trained tasks. These results confirm previous findings of training-induced improvements in non-trained WM tasks including the ability to follow instructions, but extend previous findings by showing improvements also for arithmetic. This is encouraging regarding the potential role of cognitive training for education, but it is desirable to find paradigms that would enhance the effect of the training on mathematics. One of the future challenges for studying training effects is combining large sample sizes with high quality and compliance, to detect relevant but smaller effects of cognitive training.

S. Bergman-Nutley · T. Klingberg (✉)
Department of Neuroscience, Karolinska Institutet,
171 76 Stockholm, Sweden
e-mail: torkel.klingberg@ki.se

S. Bergman-Nutley
Pearson Assessment/Cogmed Stockholm, Stockholm, Sweden

Introduction

Mathematical underachievement is estimated to affect 3–13 % of school age children (Gross-Tsur, Manor, & Shalev, 1996; Shalev, 2004) and is associated with both academic underperformance and higher risk for unemployment (Butterworth, Varma, & Laurillard, 2011), resulting in large costs to the society (OECD, 2010).

Mathematical proficiency does not only depend on specific knowledge and training in mathematics, but also more general cognitive abilities, including non-verbal reasoning (Alloway & Alloway, 2010; Geary, 2011), speed of processing (Geary, 2011) and working memory (WM) capacity (Bull, Espy, & Wiebe, 2008; Dumontheil & Klingberg, 2011; Gathercole & Pickering, 2003; Geary, 2011). WM is impaired in subjects with dyscalculia (Landerl, Fussenegger, Moll, & Willburger, 2009; McLean & Hitch, 1999; Rotzer et al., 2009), but is also correlated with mathematical performance in the general population (Bull et al., 2008; Gathercole & Pickering, 2003; Geary, 2011). Performance on WM tests is also predictive of future mathematical performance (Bull et al., 2008; Dumontheil & Klingberg, 2011; Gathercole, Pickering, Knight, & Stegmann, 2003), independent of non-verbal reasoning performance (Alloway & Alloway, 2010). The link between mathematics and visuospatial WM might be mediated by the intraparietal cortex (Rotzer et al., 2009). In

Published online: 27 September 2014

particular, brain activity in the parietal cortex during performance on a WM task is predictive of future mathematical performance in children (Dumontheil & Klingberg, 2011). The interpretation of this behavioral association and common neural substrate is not entirely clear. One link might be that relevant information needs to be kept in WM during mental operations. Another reason for the association might be the reliance on spatial coding in visuospatial WM, spatially selective attention (Klingberg, 2012) and arithmetic (Hubbard, Piazza, Pinel, & Dehaene, 2005; Knops, Thirion, Hubbard, Michel, & Dehaene, 2009).

The association of mathematical performance to WM is of special interest since a large body of evidence shows that WM can be improved by training (Chein & Morrison, 2010; Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Klingberg et al., 2005; Klingberg, Forssberg, & Westerberg, 2002). For reviews see (Klingberg, 2010, 2012). Studies investigating the effects of WM training on mathematics have thus far presented mixed results regarding such transfer (Gray et al., 2012; Dunning, Holmes, & Gathercole, 2013; Holmes & Gathercole, 2013). In one study, improved mathematical performance was not evident directly after training. But at the 6-month follow-up, the training group performed higher than at pretest (Holmes, Gathercole, & Dunning, 2009). However, there was no direct comparison to the control group at follow-up, and this study was therefore inconclusive with regard to the improvements in mathematics. In another study, adolescents with severe learning disabilities either trained on WM or a mathematical training program (Academy of Maths), and there was no significant difference in mathematical performance between the groups as a result of training (Gray et al., 2012). However, there was no third group that neither trained on WM nor mathematics and it is, therefore, impossible to know whether both or none of the groups improved.

The inconsistent results of WM training on mathematics could be due to: (1) a true lack of effect or that only certain aspects of mathematics are affected; (2) that effect occurs not directly after training but later, as a result of improved WM capacity in combination with instruction; or (3) that the effect size is small, and the existing studies include too few subjects to detect a significant effect.

In the present study, we addressed the latter of these three hypotheses by investigating the effect of WM training in a large sample of subjects in a wide age range, compared to a large control group. We measured transfer to a non-trained visuospatial WM task, a task requiring WM for instructions as well as a test of mathematics. We chose a test of arithmetic because of the suggested spatial aspect of arithmetic, spatially selective attention and visuospatial WM (Adams & Hitch, 1997; Hubbard et al., 2005; Knops et al., 2009). Furthermore, since testing was performed on a computer, with children of different ages, we chose a speeded arithmetic task that could be performed by children of all ages without ceiling effects. Given the correlation between WM capacity and arithmetic performance, we expected an effect of training, but also that it would be smaller than for WM tasks, since arithmetic depends on other factors in addition to WM.

The transfer was measured five times during the 5 weeks of training. This was done to specify the time course of the transfer. Previous studies have suggested that transfer is linearly dependent on the total amount of time spent training (Bergman-Nutley et al., 2011) as well as developing linearly during training (Jaeggi et al., 2008). This, however, might depend on the specific transfer task studied, and characterization of time might be informative for the design of future studies.

Materials and methods

Participants

The control group was recruited class-wise through an email to schools all over Sweden who were signed up for Cogmed newsletters. The teachers signed up their classes and if at least 80 % of the class performed the tasks at all five time points, the class would receive a contribution to their class trip (around 250 EU). One class of around 25 children per age group was recruited with a few additional classes in certain age groups ($n = 304$; age: M (SD) = 11.01 (2.2), age range 7–15). The data collected included age and gender, but it was not possible to connect data to a specific individual. Compliance to the test procedure was around 90 % (total that completed five sessions: $n = 275$).

The training group was recruited through clinicians in the Cogmed network who were asked to pilot the transfer measures to their end users. Subjects were included for training based on the subjective problems of inattention and working memory. A majority of subjects were diagnosed with ADHD, although this diagnosis was not verified as part of this study. In a sample of 70 subjects from this population, parents evaluated symptoms using the Disruptive Disorder Behavioral Checklist (Roberta Tsukahar, personal communication). This suggested that the children mainly had inattentive problems (score of 16) and minor problems with hyperactivity (score of 8) and ODD (score of 6). The data collected included age and gender, but it was not possible to connect data to a specific individual. All children who trained in Cogmed Working Memory Training (CWMT) with these clinicians during the summer of 2012 between the ages of 7 and 15 years, and for which information about age and gender was available, were included in the study ($n = 176$, age: M (SD) = 11.1 (2.4). Compliance to the program was very high (days trained M (SD) = 24.89 (0.71) and 88 % completed all five test sessions ($n = 155$).

Tests

The three transfer tests were: (1) a WM task called the "Odd One Out" (based on the Automated Working Memory Assessment, 2007); (2) a digital version of the "Following Instruction" task based on a previously described analog task (Gathercole, Durling, Evans, Jeffcock, & Stone, 2008) and (3), a speeded test of arithmetic developed by Cogmed, Pearson Assessment. Tasks were administered on a computer. All responses were made with the computer mouse. Only the mathematics test was timed.

Odd One Out (OOO)

In the OOO task, the participant is prompted to identify which shape out of three is the odd one and then remember its location. The stimuli are presented until the subject makes a response. The procedure is then repeated with three new shapes after which three empty slots appear for the participant to respond in the correct order of appearance, *where* the odd shapes had appeared. The task has the same progression, stopping and scoring rules as described for the Following Instructions task.

Following instructions (FI)

The task consists of common classroom items laid out on a table (e.g., eraser, crayon, box) and the task is to follow the verbal instructions given as accurately as possible. The instruction could for instance be "Click on the green eraser, then drag the black crayon to the yellow box", which would be a trial on span level 3 (because there are three items to remember what to do with). The presentation time depends on the length of the instruction. Once the participant has successfully completed the practice trials with one and two items, the task begins with two items. Two correct trials on each level will lead to progression to the next level where the item load is increased by one. The test is completed when two trials on the same level are incorrect. The final score is calculated based on performance on the highest level achieved, where at least one trial was passed. From the highest level score, 0.3 was subtracted for each incorrect answer on that level along with 0.15 for each incorrect answer on levels below the highest level achieved.

Mathematics test

The mathematics test was a speeded arithmetic test where the participants had to solve mental arithmetic problems (addition and subtraction) with two or three terms and a sum <20, excluding duplicate terms and numbers with 0 in them. As many problems as possible were to be answered during 1 min. The scoring was the sum of the correctly answered trials after subtracting the number of mistakes multiplied by 0.33 (so that random performance would give a score of 0). Subjects exhibiting signs of floor effects (below chance level for mathematics and below one correct on a span level of two items for the WM tasks) were excluded from further data analysis. 9 control subjects and 15 subjects from the training group in the analysis of Odd One Out were excluded.

Intervention

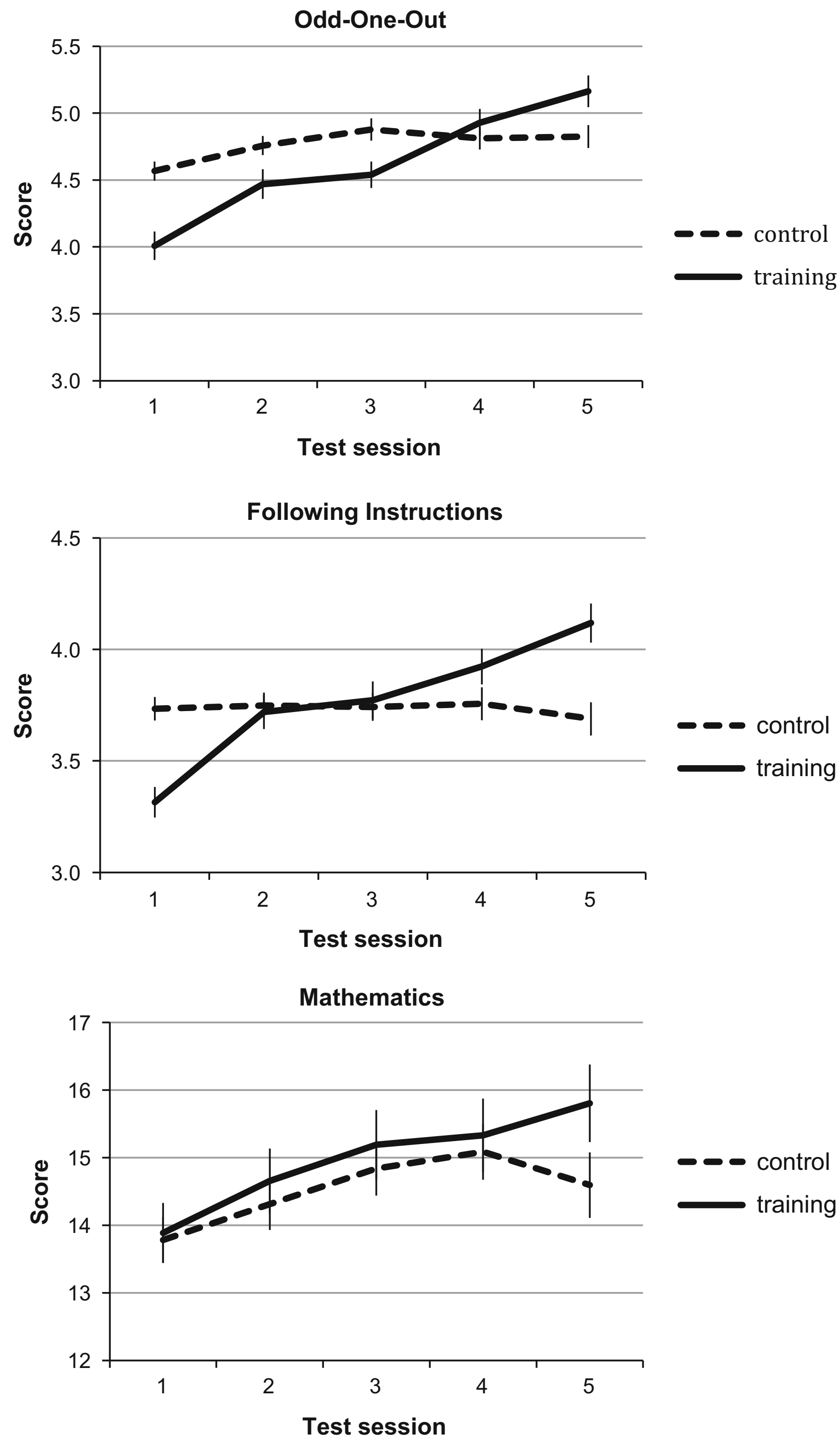
The training group went through the regular Cogmed working memory training program (described in Klingberg et al., 2005), which consists of 12 verbal and visuospatial WM tasks to be trained for 5 days/week during 5 weeks. This training program consists of 12 different WM tasks. Most have only a visuospatial content to remember, most with 2D display and one with 3D display, and three with dynamic content in which the subject recalls the positions of rotated or moving objects. Some have visuospatial and verbal elements. Tasks are changed during the training period to increase variability, so that 8 of the 12 tasks are trained in each session. Two of the tasks were present throughout the training period (a visuospatial task and a digits-backward task), whereas the other tasks were only present on parts of the training period. Therefore, these two tasks were used to measure the improvement on the trained tasks. Difficulty is dynamically adapted according to a built-in algorithm that takes an individual's previous performance into consideration. This allows training to be performed at a level that is close to, or above, the capacity limit for each individual.

Each training day corresponds to about 35 min of effective training (excluding breaks). The training is computerized and adaptive in the sense that it exposes the trainee to trials close to their capacity limit and increases the load on a trial-by-trial basis connected to the performance of the participant.

Test procedure

The three transfer tests were administered once per week for 5 weeks. The tests were computerized and fully automated so that the instructions were given through the training program. The control group received personal log-in information that was used at each test session. The

Fig. 1 Repeated testing of the three transfer tasks in the training and control groups (mean raw score, SEM)



testing was done on school computers in classrooms online in groups of 15–25 using individual headphones, with a teacher present the whole time. The training group performed the transfer tests as part of their regular training and the tests were integrated as part of the training software. This was done on computers in their homes or at the clinic for a few individuals.

Statistical testing

Statistical testing was performed with SPSS 20. The improvement in the two groups was compared using a univariate general linear model of latest outcome (T5) using group as fixed factor and age, and gender and baseline performance as covariates. This analysis is mathematically equivalent with an ANCOVA with age and baseline score as covariates.

Results

At baseline, there were performance differences between the groups for the OOO and FI (both $p < 0.001$), but not for Math ($p > 0.8$). There was no significant difference in age ($p > 0.8$), but a significant difference in frequency of males and females (Pearson Chi square, $p < 0.001$).

The correlation in performance at baseline between the OOO and FI was $r = 0.41$, between OOO and Math $r = 0.31$, and between FI and Math $r = 0.31$ (all $p < 0.001$). These correlations were also significant after partialing out the association to age (all $p < 0.05$).

The level of the trained tasks increased as the performance of the subjects improved. Improvement on an average of two of the trained tasks (one visuospatial and one verbal that are present during the entire training period, which in the program are referred to as "Index") from an average of day 2–3 to the 2 days of maximum performance was 2.23 standard deviations, which is highly significant ($p < 0.0001$).

The mean performance for the two groups at all five testing points (T1–T5) is shown in Fig. 1. The improvements in FI were linear and with minimal test–retest effects in the control group. In the OOO and the Math, there were test–retest effects apparent both at T2 and T3, after which they leveled off. As expected, for all three transfer tests the maximal difference between the training and control groups was seen at the final testing (T5).

The main effect of WM training was evaluated with a general linear model with the performance at the last testing point (T5) as the dependent variable. These data are presented in Table 1. The independent variables were group (training vs. control), age, sex and performance at baseline (T1). The group variable was significant for all three transfer tasks: OOO ($R^2 = 0.28$, β (group variable) $= 0.20$, $p < 0.0001$), FI ($R^2 = 0.27$, β (group variable) $= 0.25$, $p < 0.0001$) and Math ($R^2 = 0.58$, β (group variable) $= 0.25$, $p < 0.0001$). This showed that the training group improved significantly more than the control group from T1 to T5 on all three transfer tasks. The interaction between group $\times$ baseline performance (T1) was not significant for any of the tasks.

In addition, we performed an analysis using repeated measures from all five time points (general linear model with repeated measures). This analysis confirmed the previous analysis, with significantly larger improvements in the training group compared to the control group (linear contrast of group $\times$ time: OOO $F(1, 424) = 37.6$, $p = 2.0 \times 10^{-9}$; FI $F(1, 425) = 46.4$, $p = 3.3 \times 10^{-11}$; Math $F(1, 388) = 13.5$, $p = 0.0004$).

Table 1 Performance at the first (T1) and last (T5) testing sessions

	Control			Training		
	Mean	SD	*n*	Mean	SD	*n*
Odd One Out T1	4.6	1.3	304	4.0	1.4	176
Follow instruc. T1	3.7	0.93	304	3.3	0.91	176
Math. T1	13.8	5.6	268	13.9	5.7	162
Odd One Out T5	4.8	1.4	273	5.2	1.5	154
Follow instruc. T5	3.7	1.2	275	4.1	1.1	153
Math. T5	14.6	7.7	253	15.8	7.0	149
Age	11.0	2.2	304	11.1	2.4	176
Sex (males)			147			115

Next, we evaluated the association between training improvement and transfer improvement within the training group. Since the difference scores on the transfer tasks (T5 − T1) generally were negatively correlated with baseline performance (T1), we removed this regression toward the mean effect by regressing the T5 score on T1 score, and the standardized residual was saved and used as a measure of improvement. Improvement in Index (=2 of the trained tasks) significantly correlated with improvement in OOO ($r = 0.20$) and FI ($r = 0.23$). Improvement in FI also correlated with improvement in Math ($r = 0.36$; all $p < 0.01$).

The effect size (Cohen's d) was calculated as $(T5_t - T1_t) - (T5_c - T1_c)/SD_{T1,pooled}$, where t = training and c = control. The standardized changes $(T1 - T2)/SD_{T1}$ for the two groups are illustrated in Fig. 2. Conventionally, an effect of 0.2 is regarded a small effect, 0.5 a medium and 0.8 a large effect size. The effect for FI was strong ($d = 0.90$), for OOO medium to strong ($d = 0.67$), and for mathematics small ($d = 0.20$). However, the test scores were age-dependent, which increased the standard deviation. Age is taken into account in the statistical analysis, but not in the calculation of the effect size. When the effect size was calculated with age-normalized scores, the effect size (Cohen's delta) for mathematics was $d = 0.39$. An alternative way to calculate the effect sizes is analyzing the change in mean scores relative to the standard deviation of the change $(T5_t - T1_t) - (T5_c - T1_c)/SD_{T2-T1,pooled}$, and the effect was then 0.60 for OOO, 0.69 for FI and 0.44 for Math.

Finally, we performed a power analysis, using the test–retest difference in mathematics from each subject in a simulation that compared control group data to training group data, using 1,000 random samplings from the data at each sample size. For each sample size, a t test was performed, and the number of significant differences was then plotted against the sample size. This suggested that a power of 80 % requires a sample size of approximately 75 subjects in each group using this particular measure.

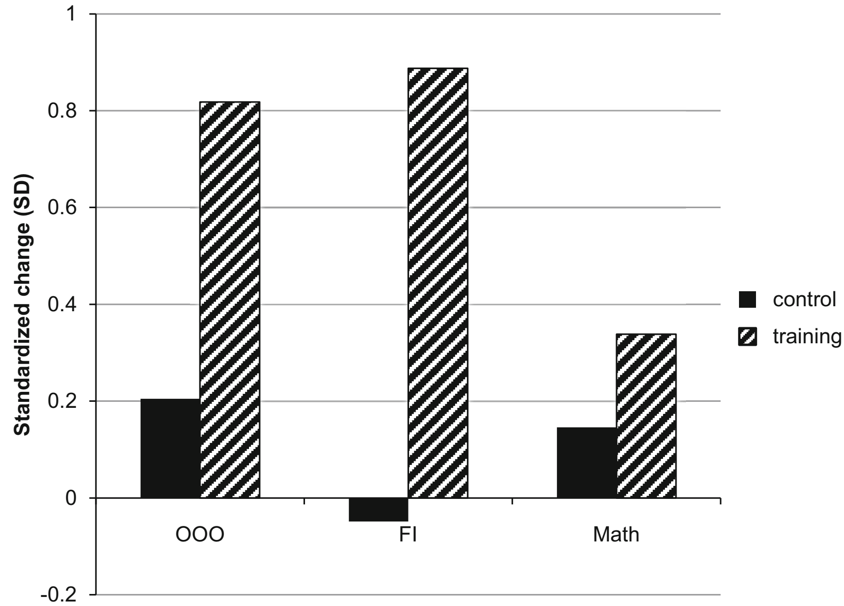


Fig. 2 Standardized change $(T5 - T1)/SD_{T1}$ for the two groups

Discussion

This study showed that WM training improved performance on transfer measures of a complex visuospatial WM task, remembering and following instructions and speeded arithmetic performance. Except for a dip at the fourth testing point in mathematics, the improvements in performance on all transfer tests increased monotonically with the amount of training, and the largest effects, as expected, at the end of training. Furthermore, the amount of improvement on the trained tasks correlated with the amount of improvement on the transfer tasks. However, the improvement in mathematics had a smaller effect size than that for the two WM tasks.

The improvement in the OOO task is consistent with previous studies using repeated baseline in children with ADHD (Holmes, Gathercole, Place, et al., 2009), as well as compared to an active control group in children with low WM (Holmes, Gathercole, & Dunning, 2009) and typically developing 4-year-old children (Bergman-Nutley et al., 2011). In the latter study, the control group performed easy problem-solving tasks with a total training time that matched that of the WM training group. Motivation was also estimated in that study and did not differ between groups. Together, these studies are consistent in showing that training induced improvement on a task, classified as a "complex working memory" task in the Automated Working Memory Assessment battery (Alloway, 2007). Although this study did not include any long-term follow-up, a previous study using the same outcome measure (Holmes, Gathercole, & Dunning, 2009) showed that the training effect remained significant at a 6-month follow-up. It is methodologically difficult to show whether these effects translate to real-world benefits, including benefits in the classroom situation. One attempt to objectively measure classroom behavior is the Restricted Academic Setting Task, in which a child is put in a classroom-like environment, with distracting toys and a less inspiring Math task to perform. The behavior of the child is then filmed and every segment rated for inattentive behavior. Using this task, Green and colleagues showed that Cogmed WM training led so significant improvement in attentive behavior (Green et al., 2012).

Improvements in ability to follow instructions have previously been demonstrated by Holmes et al. (2009) using an analog version of the FI task in children with low WM. In that study the children improved with 0.94 standard deviations, or 26 %. These data are similar to those of the current study, which showed an improvement with 0.89 standard deviations, or 24 %. This suggests that the computerized and analog versions yield similar results and the training-induced improvements are reliable also in larger groups of children. In children with low WM and attention deficits, the inability to remember instructions in everyday life is one of the most prominent features (Alloway, Gathercole, Kirkwood, & Elliott, 2009). Improved ability to retain and carry out an instruction is an ecologically valid assessment, showing that improved WM capacity is valuable in itself, independent of other transfer.

The improvement in the test of arithmetic was small, but highly significant. The increase in performance was not perfectly linear (Fig. 1), but showed a slight drop in performance for the control group at T5 for reasons that are unclear to us. There was a similar trend for the FI task. However, the analysis including all measurements in both groups in the repeated-measures analysis showed that the linear contrast (time × group) was highly significant, and thus that the effect was not due to the drop at T5, but a general trend taking all measures into account. We also performed an analysis with only the first four time points (T1–T4) in Math, and although this does not reflect the entire training period and full training effect, there was still a significant effect (linear contrast of time × group, $p < 0.05$). This improvement is consistent with the strong association between visuospatial WM and mathematics (Bull et al., 2008; Dumontheil & Klingberg, 2011; Gathercole, Brown, & Pickering, 2003; Geary, 2011), possibly based on a common neural basis in the intraparietal cortex for both tasks (Dumontheil & Klingberg, 2011; Rotzer et al., 2009). However, the improvement on the Math test, in terms of effect size, was smaller than for the two WM transfer tasks. This is hardly surprising, given that mathematical ability is dependent not only on WM capacity and top-down attention, but also on other cognitive abilities, such as non-verbal reasoning, as well as material specific knowledge and skills. In terms of change in percent improvement, the training group improved by 14 % and the control group by 6 %. Whether an improvement in mathematical performance of 0.2 standard deviations, (or 0.44 as calculated by dividing by the standard deviation of the difference score and 0.39 when using age-corrected scores), or 8 %, is valuable or worth the effort of training is an open question. As a comparison, drugs acting on the serotonin system in the brain to alleviate depression have an effect of 0.3 in severely depressed individuals (Kirsch et al., 2008) and drugs acting on the acetylcholine system of the brain, used to alleviate Alzheimer's disease, has an even lower effect (Rockwood, 2004). One study found that the effect (Cohen's *d*) of methylphenidate on visuospatial WM was about 0.2–0.3 for low-to-medium doses (0.3–0.4 mg/kg) and around 0.65 for a high dose (0.6 mg/kg) in children with ADHD (Bedard, Martinussen, Ickowicz, & Tannock, 2004). Many educational interventions have similar effects sizes (Hattie, 2009). The "What works clearinghouse" (http://ies.ed.gov/ncee/wwc/) declares an effect size of 0.25 as useful. We think that these results are encouraging and in line with theories of WM and arithmetic, but that one should make an effort to increase this effect by improving the training paradigm or increasing the length of training. One option is to combine WM training with mathematical training. Future studies could include more measures of mathematical ability as well as long-term follow-up to determine the relevance of the training effects on real-life settings, such as classroom performance.

The effect size of improvement in mathematics has important methodological implications, because a smaller effect requires a larger sample to detect a statistically significant difference. Simulations on the present set of data suggested that around 75 subjects were needed in each group to have 80 % power. Typically, training studies include 30 subjects in each group, resulting in 25 % power to detect a significant effect, and three out of four studies will fail to detect the effect (type II error). One illustration of this is the recent study by Dunning et al. (2013) on the effect of working memory training on mathematics. Using two different measures of mathematics (mathematical reasoning and number operations), the authors found effect sizes of 0.2 and 0.4, which are similar to that found in the present article, but not significant due to a lower number of subjects: they had 64 subjects, but would have needed around 150. The problem of statistical power is well known and has led to increasing sample sizes in fields such as pharmacological and genetic research. The relatively new field of cognitive training still suffers from most studies being underpowered to detect small- to medium-sized effects, which are potentially relevant. Although larger training studies have been conducted (Ball et al., 2002; Owen et al., 2010), it is a challenge to ensure quality and compliance if training is not supervised, and if training is supervised it requires large resources.

One limitation of the present study is that it used a passive control group of typically developing children, while the children in the training group had some impairment of WM. As has been argued previously (Klingberg, 2010), the ideal approach is the randomized, blinded, controlled study design, using an active control group. However, the between-group comparison included correction for baseline performance in both groups, and thus showed that the training effect was not a result of regression toward the mean, since this effect was taken into account in both groups. Furthermore, on both the OOO and FI tasks, the training group outperformed the control group at the last measure (Fig. 1). For the Math test, there was no significant baseline difference between the groups. Another potential problem with having groups from different populations is that they could differ in their ability to benefit from training. For example, severely learning disabled children have a lower learning potential compared to typically developing children (Soderqvist, Nutley, Ottersen, Grill, & Klingberg, 2012). However, genetic and behavioral evidences (Larsson, Anckarsater, Rastam, Chang, & Lichtenstein, 2012) suggest that children with attention deficits or ADHD are not a separate group, but the end of a normal distribution in attentive ability. In the present study, we did not find any interaction between baseline performance and training gain, suggesting that the gain was similar in all parts of the ability distribution.

A passive control group, as used in the present study, does not control for the effects of expectancy, which could have an effect on cognitive performance (Oken et al., 2008). However, the ideal of a randomized controlled trial

(RCT) does not mean that any other types of scientific studies do not provide important and useful information (Green, Strobach, & Schubert, 2014). In particular, the current design allowed us to include a larger sample than an RCT study would allow and to make power calculations based on actual data. Furthermore, the effect sizes found with these repeated measures of the OOO and FI were very similar to those reported in previously published randomized, controlled trials with active control groups. This suggests that the effect we found on the test of mathematics also will be very similar when similar studies are performed using randomized, controlled trials. The results from the present study could inform such future RCT studies regarding statistical power.

In conclusion, we found that WM training, using the method described previously (Klingberg et al., 2005; Klingberg, 2010), resulted in improved performance on a non-trained complex WM task, following instructions as well as arithmetic. Improved ability to remember and carry out instructions and plans is an important part of a child's everyday life, and the results of this study are thus an important confirmation of this effect. Underperformance in mathematics is also a large problem, with consequences both for the individual society. We find these results promising and encouraging for further research. Future studies could further investigate the effect on other aspects of mathematics, long-term follow-up and measures of academic performance as well as investigate the interaction between cognitive training and mathematically specific training in different ages and other approaches to enhance the effect of cognitive training on mathematics.

Conflict of interest S.B.-N. is an employee of Pearson/Cogmed and had a part-time postdoc position at Karolinska Institute providing support during the analysis of this manuscript, but has no other financial disclosures. T.K. does not have any financial association to Pearson/Cogmed.

References

Adams, J. W., & Hitch, G. J. (1997). Working memory and children's mental addition. *Journal of Experimental Child Psychology, 67*(1), 21–38. doi:10.1006/jecp.1997.2397.

Alloway, T. P. (2007). *Automated working memory assessment manual*. Oxford: Harcourt.

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106*(1), 20–29.

Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2009). The cognitive and behavioral characteristics of children with low working memory. *Child Development, 80*(2), 606–621.

Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., et al. (2002). Effects of cognitive training interventions with older adults: a randomized controlled trial. *JAMA, 288*(18), 2271–2281.

Bedard, A. C., Martinussen, R., Ickowicz, A., & Tannock, R. (2004). Methylphenidate improves visual-spatial memory in children with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 43*(3), 260–268.

Bergman-Nutley, S., Sîderqvist, S., Bryde, S., Thorell, L. B., Humphreys, K., & Klingberg, T. (2011). Gains in fluid intelligence after training non-verbal reasoning in 4-year-old children: a controlled, randomized study. *Developmental Science, 14*(3), 591–601.

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*(3), 205–228. doi:10.1080/87565640801982312.

Butterworth, B., Varma, S., & Laurillard, D. (2011). Dyscalculia: from brain to education. *Science, 332*(6033), 1049–1053. doi:10.1126/science.1201536.

Chein, J. M., & Morrison, A. B. (2010). Expanding the mind's workspace: training and transfer effects with a complex working memory span task. *Psychonomic Bulletin and Review, 17*(2), 193–199.

Dumontheil, I., & Klingberg, T. (2011). Brain activity during a visuospatial working memory task predicts arithmetical performance 2 years later. *Cerebral Cortex*,. doi:10.1093/cercor/bhr175.

Dunning, D. L., Holmes, J., & Gathercole, S. E. (2013). Does working memory training lead to generalized improvements in children with low working memory? A randomized controlled trial. *Developmental Science, 16*(6), 915–925. doi:10.1111/desc.12068.

Gathercole, S. E., Brown, L. H., & Pickering, S. J. (2003a). Working memory assessments at school entry as longitudinal predictors of National Curriculum attainment levels. *Educational Psychology, 70*, 177–194.

Gathercole, S. E., Durling, E., Evans, M., Jeffcock, S., & Stone, S. (2008). Working memory abilities and children's performance in laboratory analogues of classroom activities. *Applied Cognitive Psychology, 22*, 1019–1037.

Gathercole, S. E., & Pickering, S. J. (2003). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Educational Psychology, 70*, 177–194.

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2003b). Working memory skills and educational attainment: evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology, 18*, 1–16.

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Developmental Psychology, 47*(6), 1539–1552. doi:10.1037/a0025510.

Gray, S. A., Chaban, P., Martinussen, R., Goldberg, R., Gotlieb, H., Kronitz, R., et al. (2012). Effects of a computerized working memory training program on working memory, attention, and academics in adolescents with severe LD and comorbid ADHD: a randomized controlled trial. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 53*(12), 1277–1284. doi:10.1111/j.1469-7610.2012.02592.x.

Green, C. T., Long, D. L., Green, D., Iosif, A. M., Dixon, J. F., Miller, M. A., et al. (2012). Will working memory training generalize to improve off-task behavior in children with attention-deficit/hyperactivity disorder? *Neurotherapeutics*,. doi:10.1007/s13311-012-0124-y.

Green, Strobach, & Schubert. (2014). On methodological standards in Training and Transfer effects. *Psychological Research*.

Gross-Tsur, V., Manor, O., & Shalev, R. S. (1996). Developmental dyscalculia: prevalence and demographic features. *Developmental Medicine and Child Neurology, 38*(1), 25–33.

Hattie, J. (2009). *Visible learning*. New York: Routledge.

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009a). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science, 12*(4), F9–F15.

Holmes, J., Gathercole, S. E., Place, M., Hilton, K. A., Dunning, D. L. (2009b). Working memory deficits can be overcome: impacts of training and medication on working memory in children with ADHD. *Applied Cognitive Psychology, 12*(4), 9–15.

Holmes, J., Gathercole, S. (2013). Taking working memory training from the laboratory into schools. *Educational Psychology*. doi:10.1080/01443410.2013.797338

Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience, 6*(6), 435–448. doi:10.1038/nrn1684.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America, 105*(19), 6829–6833.

Kirsch, I., Deacon, B. J., Huedo-Medina, T. B., Scoboria, A., Moore, T. J., & Johnson, B. T. (2008). Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Medicine, 5*(2), e45. doi:10.1371/journal.pmed.0050045.

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences, 14*(7), 317–324.

Klingberg, T. (2012). Training of working memory and attention. In M. Posner (Ed.), *Cognitive neuroscience of attention* (pp. 475–486). New York: The Guilford Press.

Klingberg, T., Fernell, E., Olesen, P., Johnson, M., Gustafsson, P., Dahlstrîm, 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*(2), 177–186.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Increased brain activity in frontal and parietal cortex underlies the development of visuo-spatial working memory capacity during childhood. *Journal of Cognitive Neuroscience, 14*(1), 1–10.

Knops, A., Thirion, B., Hubbard, E. M., Michel, V., & Dehaene, S. (2009). Recruitment of an area involved in eye movements during mental arithmetic. *Science, 324*(5934), 1583–1585. doi:10.1126/science.1171599.

Landerl, K., Fussenegger, B., Moll, K., & Willburger, E. (2009). Dyslexia and dyscalculia: two learning disorders with different cognitive profiles. *Journal of Experimental Child Psychology, 103*(3), 309–324.

Larsson, H., Anckarsater, H., Rastam, M., Chang, Z., & Lichtenstein, P. (2012). Childhood attention-deficit hyperactivity disorder as an extreme of a continuous trait: a quantitative genetic study of 8,500 twin pairs. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 53*(1), 73–80. doi:10.1111/j.1469-7610.2011.02467.x.

McLean, J. F., & Hitch, G. J. (1999). Working memory impairments in children with specific arithmetic learning difficulties. *Journal of Experimental Child Psychology, 74*(3), 240–260. doi:10.1006/jecp.1999.2516.

OECD. (2010). *The high cost of low educational performance: The long-run economic impact of improving educational outcomes*. Paris: OECD.

Oken, B. S., Flegal, K., Zajdel, D., Kishiyama, S., Haas, M., & Peters, D. (2008). Expectancy effect: impact of pill administration on cognitive performance in healthy seniors. *Journal of Clinical and Experimental Neuropsychology, 30*(1), 7–17.

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., et al. (2010). Putting brain training to the test. *Nature, 465*, 775–778.

Rockwood, K. (2004). Size of the treatment effect on cognition of cholinesterase inhibition in Alzheimer's disease. *Journal of Neurology, Neurosurgery and Psychiatry, 75*, 677–685.

Rotzer, S., Loenneker, T., Kucian, K., Martin, E., Klaver, P., & von, A. M. (2009). Dysfunctional neural network of spatial working memory contributes to developmental dyscalculia. *Neuropsychologia, 47*(13), 2859–2865.

Shalev, R. S. (2004). Developmental dyscalculia. *Journal of Child Neurology, 19*(10), 765–771.

Soderqvist, S., Nutley, S. B., Ottersen, J., Grill, K. M., & Klingberg, T. (2012). Computerized training of non-verbal reasoning and working memory in children with intellectual disability. *Frontiers in Human Neuroscience, 6*, 271. doi:10.3389/fnhum.2012.00271.