Learning and Individual Differences 21 (2011) 655–661

Contents lists available at ScienceDirect

Learning and Individual Differences

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

Updating Working Memory and arithmetical attainment in school

Teresa Iuculano *, Raffaella Moro, Brian Butterworth

Institute of Cognitive Neuroscience, University College London, Alexandra House, 17 Queen Square, London WC1N 3AR, United Kingdom

ARTICLE INFO

Article history:
Received 26 April 2010
Received in revised form 18 November 2010
Accepted 10 December 2010

Keywords:
Working-memory
Updating
Arithmetical attainment

ABSTRACT

Here we wished to determine how the sub-components of Working Memory (Phonological-Loop and Central Executive) influence children's arithmetical development. Specifically, we aimed at distinguishing between Working Memory inhibition and updating processes within the Central Executive, and the domain-specificity (words and numbers) of both subcomponents in a population of children with low attainment in arithmetic and their age matched typically-attaining controls. We show that both groups were similar for phonological loop abilities, while Working Memory updating demonstrated a domain-specific modulation related to the level of children's arithmetical performance. Moreover, inhibition processes interacted with domain-specificity and arithmetical attainment. These results are particularly relevant to the diagnostic assessment of arithmetical ability and should be considered in existing tests of arithmetical development.

1. Introduction

1.1. Working Memory and arithmetic

It is widely assumed that Working Memory (WM) is important in calculation (Baddeley & Hitch, 1977; Hitch, 1978), and in particular, during the early development of arithmetical skills (e.g. Gathercole & Pickering, 2000; Geary, 1990, 1993; Geary & Hoard, 2001; Ginsburg, 1997; Jordan & Montani, 1997; Kirby & Becker, 1988; Russell & Ginsburg, 1984; Shalev & Gross-Tsur, 2001). Support for this comes from two sources. First, there are tasks assessing the effects on calculation in arithmetically competent adults when the WM system is disturbed by experimental manipulations (e.g. De Rammelaere, Stuyven, & Vandierendonck, 1999, 2001; Furst & Hitch, 2000; Lemaire, Abdi, & Fayol, 1996; Logie, Gilhooly, & Wynn, 1994). The second relates individual differences in WM to individual differences in arithmetical attainment in school (e.g. Gathercole & Pickering, 2000; Geary, 1990, 1993). A key distinction made in studies investigating the relationship between WM and arithmetical abilities has focused on the role of the different subsystems of WM as originally proposed by Baddeley and Hitch (1977). A primary distinction has been made between the Phonological Loop (PL) and the visuo-spatial sketchpad (VSSP). The former has been associated with solving single-digit addition problems (Hecht, 2002; Seyler, Kirk, & Ashcraft, 2003); while the latter has been linked with the encoding of visually presented problems (Logie et al., 1994). Yet, the third component, the Central Executive (CE) system has been thought to play a key role in aspects of calculation that require the storage and manipulation of intermediate results online, by updating the results of operations such as carrying and borrowing. In this model, the CE was originally thought of as the system 'to which all the complex issues that did not seem to be […] specifically related to the two subsystems were assigned' (Baddeley, 2003).

Given the distinctions proposed by the model, it has become important to draw a distinction between the different components of WM and their relationship with arithmetical abilities. This notion has lead to mixed results on the role of the different subsystems in supporting calculation. The key distinction made in both types of study is between Central Executive (CE) processes and processes dependent on the Phonological Loop (PL) subsystem of the WM model proposed by Baddeley (1986). Maintenance processes are intuitively plausible as the principal locus of the WM contribution to calculation, since intermediate results from operations such as carrying and borrowing are required by mental computation. However, where CE and PL can be experimentally distinguished, it is CE that seems more critical for calculation. For example, De Rammelaere et al. (1999, 2001), found that articulatory suppression, which should interfere with PL but not with CE did not affect calculation; while random interval generation, thought to be the responsibility of CE, did reduce arithmetical performance. Furthermore, it has been found that children with specific difficulties in arithmetic (compared to both age-matched and ability-matched controls) do not differ on tasks that rely primarily on PL, such as immediate serial recall (e.g. digit span), but perform worse on tasks tapping CE (McLean & Hitch, 1999; Passolunghi & Siegel, 2001; Siegel & Ryan, 1989). On the other hand, it is the PL component of Working Memory which has specifically been associated with arithmetical impairments (Hecht, Torgensen, Wagner, & Rashotte, 2001; Hitch &

Abbreviations: CE, Central Executive; FS.I.Q., Full Scale Intelligence Quotient; HI, High Inhibition; LA, Low Arithmetic; LI, Low Inhibition; PL, Phonological Loop; TA, Typical Arithmetic; WM, Working-Memory; WML, Working-Memory Load.

* Corresponding author. Tel.: +44 20 7679 5524.
E-mail address: t.iuculano@ucl.ac.uk (T. Iuculano).

1041-6080/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.lindif.2010.12.002

McCauley, 1991; McLean & Hitch, 1999; Swanson & Sachse-Lee, 2001). Such a notion has been supported by dyscalculic pupils having shorter span performance than controls (Koontz & Berch, 1996). Indeed, both components of the WM system could play a role in the typical or atypical development of numerical cognition, but at distinct stages of development. PL could therefore contribute to the development of the number vocabulary and counting strings, while the CE could be responsible for tracking mental operations (Noël, Seron, & Trovarelli, 2004).

The most convincing evidence relating WM to mathematical achievement in children is a longitudinal study investigating children from kindergarten to third grade by Geary et al. (2009). They used the Working Memory Test Battery for Children (Pickering & Gathercole, 2001) to separately measure CE, PL and the visuospatial sketchpad (VSSP). This study has found that different systems of WM can discriminate between different levels of mathematical impairments in children as well as their different levels of math proficiency (Geary et al., 2009). Specifically, Geary and colleagues underline the importance of the CE component for the ability of correctly retrieving simple addition facts, which is of particular relevance to the present study.

Now it is generally assumed that WM is a domain-general system that supports a wide range of cognitive processes. Yet, it has also been suggested that in fact there is a domain-specific coding. In particular, that numbers may be maintained differently or separately from words (Butterworth, Cipolotti, & Warrington, 1996). It may be important therefore to distinguish number from word memory, and to directly address the issue of selecting and maintaining task-relevant information.

In fact, in a cross-sectional study by Geary et al. (2009), it was found that the best predictors for inclusion of Mathematics Learning Disability (MLD) and Low Achieving (LA) were purely numerical — a number line task and the ability to discriminate the numerosity of sets, which suggests that “these children have a poor number sense, in support of other findings” (Butterworth & Reigosa, 2007, p426). By contrast, WM measures in this cross-sectional study gave rather mixed results. “The best constellation of variables for predicting membership in the MLD class included IQ and the phonological loop and visuo-spatial sketch pad scores, the means for all of which were about 1 SD below average” with the PL scores being higher for the LA than the typically-achieving group. However, the CE variable did not emerge as a significant predictor of MLD class membership in their analyses contrary to previous findings by Geary, Hoard, Byrd-Craven, Nugent, and Numtee (2007).

One possible reason for the mixed findings on measures of CE and their relationship to mathematical abilities may be the particular paradigm used. The tasks tapping the CE components of WM involved counting and counting errors in a separate test were found as good predictors of the CE measures. Thus, to better evaluate the role of CE, it is important to have a measure that directly taps the critical function of CE in arithmetic — selecting and maintaining task-relevant numerical information. Of particular relevance to the current study, are processes of inhibiting and updating the contents of memory (Miyake et al., 2000). The “updating function goes beyond the simple maintenance of task-relevant information in its requirement to dynamically manipulate the contents of working memory” (Miyake et al., 2000, p.57).

A recent neuroimaging study on healthy adults claims for a special role for numbers in the PL component of WM (Knops, Nuerk, Fimm, Vohn, & Willmes, 2006). Importantly this study demonstrates stimulus-specific modulation of numerical stimuli in the intraparietal sulcus, a region often associated with numerical processing (see Dehaene, Piazza, Pinel, & Cohen, 2003), but not for word stimuli. The authors highlight the importance of considering the a-verbal semantic component intrinsic to number stimuli when designing WM paradigms for neuropsychological testing.

Taken together, the current evidence for a complex relationship between arithmetical abilities and WM suggests a need for specific WM paradigms that distinguish among the different subcomponents of WM and the content to be remembered.

1.2. Measuring maintenance and updating processes in Working Memory: the Updating task

The present study investigates the ability to update relevant information and the ability to inhibit irrelevant information in a group of children who presented a selective impairment in arithmetic. In addition to established tests to assess the capacity of the PL (digit and word span tests), a novel test was developed based on an Updating task previously used by Palladino, Cornoldi, De Beni, and Pazzaglia (2001). Updating of WM was defined as the amount of information recalled after being held and manipulated in WM. This concept is similar to what Broadbent (1958) called “channel capacity”, and Cowan (1995) called the “capacity of the focus of attention”. Inhibition in the CE was defined as the amount of information to be suppressed according to a prespecified criterion (see Method for details). This concept is similar to the 'selective filter' defined by Broadbent (1958) and to 'controlling the direction of attentional focus' as proposed by Cowan (1995).

The advantage of using this Updating task is that it can pose variable demands on maintenance and inhibition processes separately (Moro, 2008). The task requires participants to recall information that is relevant according to a given criterion and at the same time inhibiting irrelevant information.

1.3. Aims of the study

In this study we tested a group of children exhibiting a selective impairment in exact calculation (addition) and their matched control peers in two canonical span tasks and two novel updating tasks using numbers and words as stimuli. Our first aim was to determine whether selective impairments in arithmetical abilities could be attributed to a deficit in WM. Moreover, we wanted to investigate which subcomponent of WM might be impaired or spared in this population of children. Specifically, whether impairments could be seen in maintenance processes (PL subcomponent) as measured by the span tasks, or in updating and inhibition processes (CE subcomponent). Furthermore, by manipulating WM load and inhibition levels in the updating tasks, we aimed at better differentiating the updating and inhibition processes of the CE component of WM. Finally, we aimed to determine whether different stimulus categories could discriminate between the two groups of children's WM abilities.

2. Method

2.1. Participants

Participants (8 to 9 year olds) were recruited from three different State Middle Schools in the London area and assigned to two groups: Low Arithmetic group (LA, n = 11) and Typical Arithmetic group (TA, n = 22).

2.1.1. Participant selection procedure

Participants in the experimental group (LA) were initially selected upon teachers' assessment: teachers were asked to nominate children who they felt were of normal intelligence but had serious difficulties during Numeracy lessons, and the final selection followed additional testing (*see below*). Participants in the control group (TA) comprised children who were of the same gender and from the same class with the children in the experimental group, to minimise the effects of instruction. Prior to participation in the study participant assents and parental consents were collected for each child. This study was approved by the UCL Ethics Committee.

Participant screening included a battery of mathematics and I.Q. assessments. Each child was tested individually in a quiet room of the schools in four different sessions (each one lasting no longer than 20 min). Mathematics assessments consisted of a standardised software used for the diagnosis of mathematical difficulties on the basis of age norms (*Dyscalculia Screener*, Butterworth, 2003; see also Landerl, Bevan, & Butterworth, 2004). This software comprises three computer-controlled, item-timed math tests plus one simple reaction time test to assess whether slowness in responding to the numerical tasks was due to general slow reaction times. The three math tests are divided into two subscales: 'Capacity subscale', which involves a dot enumeration task and a number comparison task; and an 'Achievement subscale', which involves an addition task. The 'Capacity subscale' is designed to assess basic numerical capacities. Defective performance on this subscale defines developmental dyscalculia (Butterworth, 2003). Children with dyscalculia were not included in this study. The 'Achievement subscale' includes tasks which depend also on the learning experience of the pupil. The software computes a combined measure of accuracy and reaction times – inverse efficiency – by dividing the adjusted median reaction times by the proportion of correct responses for each of the three tasks. The test average on accuracy of the nationally standardised score is 100 (15 SD).

I.Q. was examined only in the experimental group using the WISC-III full protocol (Wechsler, 1996). Results were then pro-rated for the arithmetic subtest. All participants obtained an average I.Q. score according to their age group (mean FS.I.Q. = 102.7, SD = 15.47) except one, who was discarded from subsequent analyses (FS.I.Q. = 67).

Low Arithmetical ability was defined by the following criteria: 1) a standardised score below 81 on the 'Achievement subscale' of the Dyscalculia Screener, which is the equivalent of the bottom 7% of the population; 2) a performance within the normal range (greater than 89) in both tasks of the 'Capacity subscale' in order to exclude dyscalculic participants; and 3), an I.Q. score within the normal-range for their age group (FS.I.Q. score > 78). Thus, the LA group was defined as being in the bottom two stanines[1] for their age-group for timed addition, but was not dyscalculic according to the results on the two capacity tasks. Participants in the Typical Arithmetic group (TA) displayed average performance on the two capacity tasks of the Dyscalculia Screener as well as on the achievement task (fourth stanine or above).

Further statistical analyses on the two groups of interest revealed that the TA group was significantly better than LA in addition accuracy (87.3% (SD 8.47) vs 58.9% (SD 7.85), $p<0.01$).

2.2. Mathematics assessments

2.2.1. Dyscalculia Screener

2.2.1.1. Simple reaction time task. The child is asked to press a key as soon as a black spot is presented (in random locations and after randomised intervals) on a white computer screen. Reaction times on the three numerical tests are adjusted to take this measure into account.

2.2.1.2. Dot enumeration task. This task is one of the measures of the 'Capacity subscale'. The child is asked to decide, using a key press response, whether the numerosity of a random array of dots presented on a left panel of the screen matches the numerosity of a digit in the right panel. The range of numerosities used was between one and ten. Accuracy and speed were emphasised. This task includes sixty-eight trials.

[1] Please note that *Stanine* (STAndard NINE) is a method of scaling test scores on a nine-point standard scale with a mean of five (5) and a standard deviation of two (2), which is now widely used in educational assessments (e.g. Canada). The second Stanine corresponds to the bottom 7% of the population.

2.2.1.3. Number comparison task. This task is part of the 'Capacity subscale'. The child is presented with two Arabic single digits – one on the left and one on the right side of the computer screen – and is asked to indicate the larger number in numerical value by a key press. Digit pairs used were between one and nine (excluding number five). The software registers both accuracy and reaction time for each of the forty-two trials.

2.2.1.4. Addition task. The addition task is part of the 'Achievement subscale'. The child is asked to verify the result of single-digit addition problems presented in the middle of the screen and press a key according to whether the result is correct or not for a total of twenty-eight trials. Accuracy and speed are emphasised.

2.2.2. Approximate arithmetic

It has been argued that an innate system for processing approximate numerosities forms the basis of arithmetical development (Barth et al., 2006; Gilmore, McCarthy, & Spelke, 2007; Halberda, Mazzocco, & Feigenson, 2008; but see Iuculano, Tang, Hall, & Butterworth, 2008). We therefore exploited a set of tasks which used non-symbolic stimuli requiring only a grasp of approximate numerosity, and compared performance of the LA group to the age-matched TA group. These assessments were programmed in Matlab 6.5 using the Cogent application.

2.2.2.1. Approximate numerosity tasks. Three approximate tasks (see Iuculano et al., 2008) were used to assess approximate non-symbolic arithmetic and numerical comparison abilities. Please note that this battery of tasks is not part of a standardised test for arithmetic, but it is rather intended as an experimental set of tasks comparing the two groups of interest (see Table 1).

In the Comparison task a set of blue dots appeared on the upper left side of the computer screen (1300 ms) and then moved behind a black box placed on the lower left side of the screen (650 ms). An array of red dots subsequently appeared on the upper right side of the screen (1300 ms) and moved down (650 ms) to the bottom right of the screen. The child had to select the array with more dots.

In the Addition task, one array of blue dots appeared on the upper left side of the computer screen (1300 ms) and moved down behind a black box placed on the lower left side of the computer screen (650 ms). Another array of blue dots would then appear in the same position as the previous one (1300 s) and moved down behind the same black box (650 ms). Finally, as in the comparison task, an array of red dots appeared on the upper right side of the screen (1300 ms) and moved down (650 ms) to the bottom right of the screen. The child had to decide which had more dots, the box with two sets of blue dots or the comparison array of red dots.

Table 1
Approximate task scores for LA and TA groups.

Approximate tasks	LA group	TA group	Statistical analyses
	Mean (SD)	Mean (SD)	
Approximate comparison			
Accuracy	60(13)	67.4(13)	$t(30) = 1.48$
RTs	703(355)	710(239)	$t(30) = .073$
Inverse efficiency	52.1(29.3)	45.9(17.6)	$t(30) = -.075$
Approximate addition			
Accuracy	76.7(8.3)	70.4(16.1)	$t(30) = -1.14$
RTs	784(532)	761(432)	$t(30) = -1.32$
Inverse efficiency	45.1(34.3)	51.8(45.9)	$t(30) = .04$
Approximate subtraction			
Accuracy	52.9(8.5)	61.7(11.8)	$t(30) = 2.11^{*}$
RTs	824(521)	699(302)	$t(30) = -.86$
Inverse efficiency	67.4(46.7)	51.6(32.8)	$t(30) = -1.1$

* $p<.05$.

In the Subtraction task, an array of blue dots appeared on the upper left side of the computer screen (1300 ms) and moved down behind a black box positioned on the lower left side of the screen (650 ms). A subset of the same array of blue dots then moved out of the box and disappeared off the screen (650 ms). Finally, as in the other two tasks, an array of red dots appeared on the upper right side of the screen (1300 ms) and moved down to the bottom right side of the screen (650 ms). The child had to decide which had more dots, the blue dots remaining in the box or the red dots.

The final result of numerosities for both the addition and the subtraction operations was equated across the two tasks. The numerosities used in the three tasks ranged between ten and fifty-eight dots. Each task comprised twenty-four trials. Accuracy and speed were both emphasised.

2.2.3. Working Memory assessments

All Working Memory assessments were administered in the fourth session of testing. The order of administrating tasks was counter-balanced among participants. The dependent variable used was accuracy which was defined in terms of the percentage of correct recalls from a given list.

2.2.3.1. Span tasks. Two verbal span tasks were used: a Digit Span and a Word Span task, both forward and backward. The stimuli used in the Digit Span forward and backward were identical to those used in the WAIS-III scale (Wechsler, 1997). Stimuli in the Word Span task were of two semantic categories – Animals and Objects – and were selected in order to match the digit stimuli in number of syllables (one to five) and word lengths (number of letters). Lists were of increasing complexity from two to nine items. Stimuli were verbally presented by the same experimenter at the rate of one item per second.

The task instructions were the following for both forward tasks (Digit and Word): "I am going to say some numbers (or words). Listen carefully, and when I am through, I want you to say them right after me. Just say what I say". Instructions for the backward tasks were: "Now I am going to say some more numbers (or words). But this time when I stop, I want you to say them backward, from the last one that I say to the first one. For example, if I say 7–1–9, what would you say?" If the child was correct, the task will begin with two practice trials, otherwise the experimenter will give the correct answer of the problem (9–1–7) and assure the child had understood the task, before administering the practice trials.

2.2.3.2. Updating tasks. A task of WM was devised to test the participants' ability to update relevant information and inhibit irrelevant information during a task of free recall with a semantic criterion (Moro, 2008). It was adapted from Palladino et al. (2001).

Sixteen lists of words (eight lists of names of animals and eight lists of names of objects) and sixteen lists of two-digit numbers (odd in half of the lists even in the other half) were presented to the children who were required to retain the relevant stimuli based on an ongoing semantic criterion (magnitude of the stimuli). The words were bi- or tri-syllabic highly familiar and imaginable nouns initially selected from a list by Burani, Barca, and Arduino (2001) and also via a pilot study in which 23 adult participants were asked to judge, on a scale from 1 (very small) to 9 (very big), the dimensions of 53 animals and 100 objects. The order of presentation was randomised. Only the items with a clearly discriminable size were used. A second pilot study (with 20 adult participants) was conducted, in order to check the discriminability of the selected items within the lists. The lists were balanced for number of syllables and length of the word (number of letters).

Stimuli in the Number Updating task were double-digit numbers ranging from 22 to 99. They were associated to the animals and objects according to the size-judgement from the pilot study (i.e. the smallest animal/object used in the list would correspond to number 22, which was the smallest number in the list of numbers). In this way the lists were similarly constructed so that each number corresponded to an object or animal. The numbers excluded were teens, multiples of ten, and numbers containing 1 as a unit: since numbers with "1" as decade (10–19) were removed, numbers with "1" as the unit (e.g. 21, 31, …, and 91) were also removed so that each digit (2–9) would have an equal rate of occurrence. Moreover, in this way there was an equal number of odd and even numbers. A possible source of confusion could have been that the number of syllables composing the double-digit number stimuli is bigger (3 to 5 syllables) than single-digit numbers, but the use of two-digit numbers was necessary in order to have a large number of different items and match the two tasks.

Four practice trials (two for each task) preceded the actual experiment to ensure that the child understood the task requirements. The stimuli were verbally presented by the experimenter at the rate of one item per second and the child was required to recall a predefined number of the smallest items presented. This procedure required the child to constantly update the incoming information and to inhibit or suppress the irrelevant items.

WM load was manipulated by having four recall conditions with participants having to recall one (WML1), two (WML2), three (WML3) or four (WML4) of the smallest items presented. From now on we will refer to this factor as Working Memory Load. *Inhibition* was a two level factor with the participant having to ignore (or initially recall and then inhibit) one item (condition of Low Inhibition, LI) or three items (condition of High Inhibition, HI). As a result, the list length varied between two and seven items. An example of list with three items to maintain and three items to inhibit, is: "giraffe–pelican–tortoise–tiger–chicken–dolphin". Here, the participant must remember the three smallest animals in the list (i.e. pelican, tortoise, and chicken) while inhibiting the recall of the other three. An example of a list with numbers, where two items had to be recalled and three had to be inhibited is: "26–68–92–66–35". Here the items to recall are 26 and 35 (i.e. the two smallest numbers in the list). In order to perform correctly, the child, while listening to the presented items, has to constantly update the information with the new item presented and to inhibit one of the previously recalled items that is no longer fulfilling the criterion.

The percentage of correct items recalled was considered as a measure of WM.

3. Results

Significant differences between the two groups on the WM assessments were assessed by repeated measures ANOVA. The model for the Span tasks had Task (Word task and Number task) and Condition (Forward and Backward) as the within subjects factors and Group (LA and TA) as the between subjects factor.

The Updating task used a $2 \times 4 \times 2 \times 2$ mixed design with Task (Word and Number), Recall (one, two, three and four items to recall) and Inhibition (Low — one item to inhibit and High — three items to inhibit) as within subjects factors. Group (LA and TA) was the between subjects factor. Accuracy was defined as the percentage of correct recalls.

Greenhouse–Geisser corrections were applied to all factors with more than two levels to remedy violations of sphericity (Keselman & Rogan, 1980).

3.1. Span tasks

Repeated measures ANOVA revealed a main effect of Task ($F(1,30) = 4.8$, $p = .036$, $\eta^2 = .13$): the Number task was significantly better than the Word task; and a main effect of Condition ($F(1,30) = 154.6$, $p < .001$, $\eta^2 = .84$): the Forward condition was significantly better maintained than the Backward condition. There was no main effect of Group ($F(1,30) = .078$, $p = .78$, $\eta^2 = .003$) (Fig. 1). All the

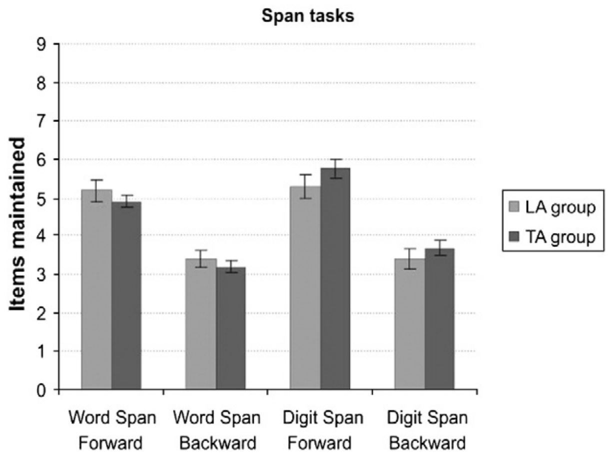

Fig. 1. Mean of maintained items for the LA group (light grey) and for the TA group (dark grey) for each of the Span tasks. Error bars indicate standard errors.

interactions were not significant: Task by Group ($F(1,30)=3.6$, $p=.068$, $\eta^2=.1$); Condition by Group ($F(1,30)=.03$, $p=.85$, $\eta^2=.001$); Task by Condition ($F(1,30)=.89$, $p=.35$, $\eta^2=.03$) and Task by Condition by Group ($F(1,30)=.29$, $p=.59$, $\eta^2=.009$).

3.2. Updating tasks

A repeated measures ANOVA revealed a main effect of Task: the Word task was more accurate than the Number task [$F(1,30)=53.87$, $p<.001$, $\eta^2=.64$]. There was a main effect of Working Memory Load [$F(2.33, 69.79)=190.799$, $p<.001$, $\eta^2=.86$] and a main effect of Inhibition [$F(1,30)=117.99$; $p<.001$, $\eta^2=.79$]: in both tasks, the proportion of correct recalls decreased with increased WM load and it was modulated by increased inhibition of irrelevant information. There was also a Task by WML interaction [$F(2.75, 82.61)=7.51$; $p<.001$, $\eta^2=.2$]: the two tasks only differed in the WML3 and WML4 conditions where the accuracy of the Number task was lower than the Word task ($p<.001$). The Task by Inhibition interaction was also significant [$F(1,30)=7.42$; $p<.05$, $\eta^2=.2$]: the two tasks only differed in the Low Inhibition condition ($p<.05$). Moreover, the Task by Inhibition by WML interaction was also significant [$F(2.74, 82.32)=3.79$; $p<.05$, $\eta^2=.11$]. Post-hoc analyses (paired sample t-tests) showed that the two tasks (Words and Numbers) differed for both levels of Inhibition only in the WML3 condition ($p<.005$)

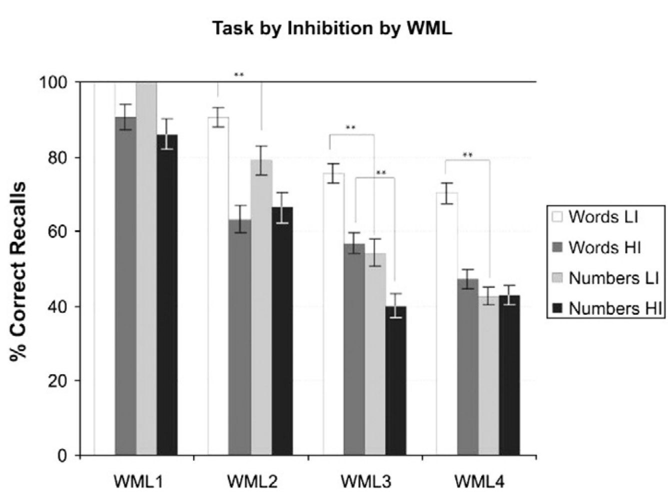

Fig. 2. Percentage of correct recalls for the four WM Load conditions (x-axis) on the two Inhibition conditions for both tasks. **$p<.005$. Error bars indicate standard errors.

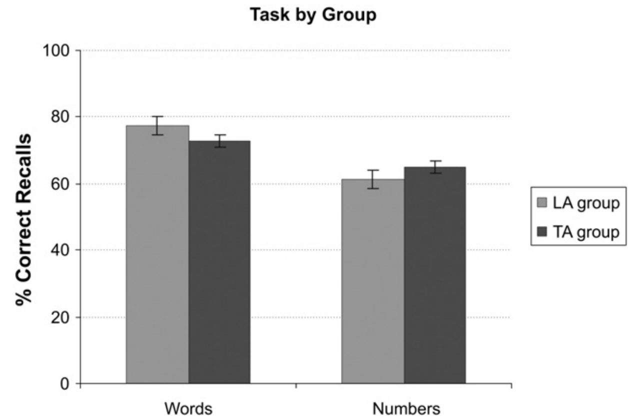

Fig. 3. Percentage of correct recalls for the Word task and the Number task in the LA and the TA groups. **$p<.005$. Error bars indicate standard errors.

(see Fig. 2). The interaction of WML by Inhibition was not significant ($F(2.73, 81.9)=2.24$, $p=.095$, $\eta^2=.07$).

There was no main effect of Group [$F(1,30)=.008$, $p=.93$, $\eta^2=.00$]. However, a Task by Group interaction was found [$F(1,30)=6.72$, $p<.05$, $\eta^2=.18$] (see Fig. 3). The three-way interaction Task by Inhibition by Group was also significant [$F(1,30)=8.34$, $p<.01$, $\eta^2=.22$] (see Fig. 4A). Post-hoc analyses (independent sample t-tests) revealed that the LA group performed significantly better than the TA group in the HI condition of the Word task ($p<.05$) (see Fig. 4B).

The other interactions were not significant: Inhibition by Group ($F(1,30)=1.7$, $p=.202$, $\eta^2=.05$); WML by Group ($F(2.33, 69.79)=.57$, $p=.64$, $\eta^2=.02$); Task by WML by Group ($F(2.75, 82.61)=1.04$, $p=.37$, $\eta^2=.03$); WML by Inhibition by Group ($F(2.73, 81.9)=.36$, $p=.78$, $\eta^2=.01$); and Task by WML by Inhibition by Group ($F(2.74, 82.32)=.093$, $p=.95$, $\eta^2=.003$).

4. Discussion

This study employed a new updating task to assess the contribution of WM components to the processes of calculation in 8–9 year olds. It showed first, that updating in the Word task was easier than updating in the Number task. Second, it showed clear and significant effects of Working Memory Load (number of items to recall) and Inhibition (number of items to inhibit). Moreover, the LA group displayed better performance for the High Inhibition of word stimuli. This is consistent with the proposal for domain- or material-specific capacities in WM (Butterworth et al., 1996; Semenza, Miceli, & Girelli, 1997). Under this notion, two-digit numbers display greater semantic and syntactic complexity compared to long nouns. This interpretation is consistent with our current findings that multi-digit number stimuli are more difficult than word stimuli in accordance with previous data in healthy adults (Moro, 2008). Furthermore, words can be remembered through verbal encoding and image encoding, while numbers cannot (Paivio, 1991; Paivio, Walsh, & Bons, 1994). This could explain why, in general, highly imaginable concrete nouns are more easily remembered than numbers and other word categories (i.e. abstract nouns or verbs). On the other hand, one digit numbers in the span tasks were found to be recalled better than nouns of the same length for both groups.

This study distinguished two groups of 8–9 year olds on the basis of their speed and accuracy on a standardised test of numerical capacity and attainment in curriculum exact arithmetic (Butterworth, 2003). Additionally, they were tested on their ability to carry out approximate addition and subtraction with non-symbolic (dot array) stimuli since these are held to be foundational for exact arithmetic (Gilmore et al., 2007; Halberda et al., 2008). Two groups were readily distinguishable on the basis of the arithmetical attainment in speed

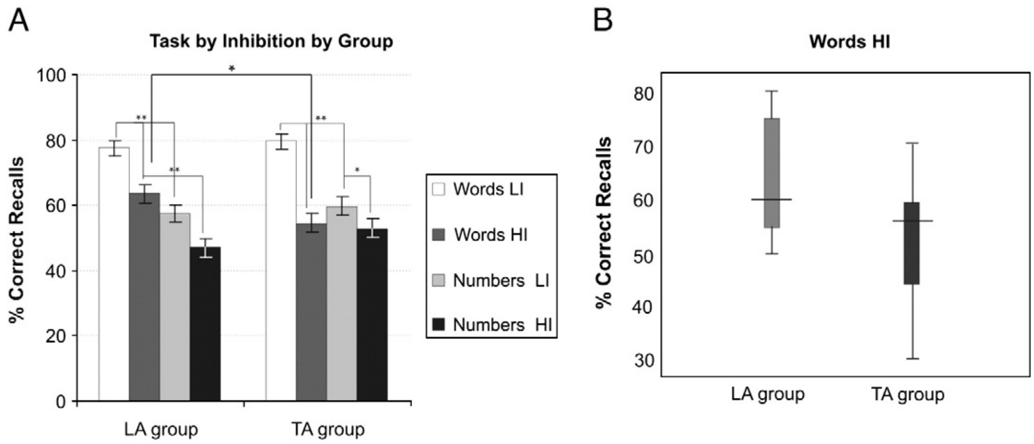

Fig. 4. A. Percentage of correct recalls for the Word and the Number tasks with levels of Inhibition in the LA and the TA groups. (*<.05; **<.005). B. The LA group performed significantly better than the TA group on the Word task at High Inhibition level. Error bars indicate standard errors.

and accuracy of exact addition, but not on their ability to carry out the capacity tasks of dot enumeration and number comparison. Moreover, the groups did not differ on any of the approximate arithmetic tasks.

Comparing the two groups (TA and LA) on measures of WM, our findings support a complex relationship between the components of WM and calculation. Not only both span tasks forward, which are robust assessments of PL, but also both span tasks backward, which also rely strongly on PL, were performed at the same level in each group, suggesting that these WM processes were not critical for distinguishing typical from low attainment in arithmetic.

As was found by McLean and Hitch (1999) among others, forward span does not predict calculation ability, and likewise in this study it does not distinguish LA from TA. Moreover, in a study of developmental dyscalculia in 8–9 year olds, Landerl et al. (2004) were able to match dyscalculics on a Digit Span task with the typical learners. Our results suggest that the immediate serial recall and rehearsal of information (either words or numbers) do not discriminate between the two groups tested and therefore do not seem to be an optimal measure in order to assess the hypothesis of a causal or correlational account between WM and calculation. However, we are not suggesting that these tasks should be removed from the existing batteries of neuropsychological assessments, rather implemented with tasks which tap the specific subsystems of WM (in this case the CE component) and which discriminate between the stimulus materials used.

In the updating tasks, we found no overall difference between the two groups. Both groups handled increasing Working Memory Load and the effects of Inhibition equally satisfactorily. Yet, a Task by Group interaction was found suggesting that the LA group performed better on Word than Number stimuli, while the contrary was true for the TA group (see Fig. 3). In our view, this finding stresses a need to specify the stimulus type in WM tasks (see also Knops et al., 2006), especially when investigating learning disabilities.

Additionally, the three-way interaction of Task × Inhibition × Group, suggests that High Inhibition had a greater effect on Words than Numbers in the TA group while it had a greater impact on Numbers than Words in the LA group. Surprisingly, the LA group performed better on Words than Numbers while no significant effect was found in the other direction. One explanation for this could be the fact that Words elicited a bigger effect on semantic processes and were a better discriminator of group performance. However, it is possible that we were unable to detect any specific differences in Number stimuli due to our relatively small sample size in the LA group. Yet, Geary et al. (2009) also found facilitation in tasks tapping the PL in Mathematical Learning Disabilities. Our study extends this evidence to the CE component of WM and highlights its relationship to the differing levels of inhibition.

To conclude, our study provides evidence for intact rehearsal and maintenance processes in WM by measures of the PL (span tasks). Additionally, it demonstrates the relationship of a domain-specific updating function within the CE component of WM as implied by studies of neurological patients (Butterworth et al., 1996; Semenza et al., 1997). Finally, our results provide evidence for the importance of considering the different aspects of CE subsystems such as Working Memory Load and Inhibition processes, as they might both be crucial for discriminating WM performance in clinical populations.

Finally, we believe that these results are of particular relevance for the assessment of relationships between WM and arithmetical abilities. We propose that these domain-specific updating tasks can be used as a basis for respective studies examining the reliability of the present approach by recruiting larger samples of LA individuals and also pupils with Developmental Dyscalculia. The ultimate goal of such studies should focus on optimizing assessments that better characterize the educational development of children with and without learning disabilities.

Acknowledgments

We would like to thank all children and their parents, who participated in the study. We would also like to thank the three schools who took part in the study and the teachers who helped in the recruiting stage of the study. We are grateful to three anonymous reviewers for their helpful comments on an earlier version of this manuscript.

References

Baddeley, A. D. (1986). *Working memory.* Oxford: Clarendon Press.

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Review Neuroscience*, *4*, 29−39.

Baddeley, A. D., & Hitch, G. J. (1977). In G. Bower (Ed.), *Commentary on "working memory". Human memory: Basic processes.* (pp. 191−197) New York: Academic Press.

Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition*, *98*(3), 199−222.

Broadbent, D. E. (1958). *Perception and communication.* London: Pergamon Press.

Burani, C., Barca, L., & Arduino, S. L. (2001). Una base di dati sui valori di eta' di acquisizione, frequenza, familiarita', immaginabilita', concretezza, e altre variabili lessicali e sub-lessicali per 626 nomi dell'italiano. *Giornale Italiano di Psicologia*, *4*, 839−854.

Butterworth, B. (2003). Dyscalculia screener. London: nferNelson Publishing Company Ltd.

Butterworth, B., Cipolotti, L., & Warrington, E. K. (1996). Short-term memory impairment and arithmetical ability. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology*, *49*, 251−262.

Butterworth, B., & Reigosa, V. (2007). Information processing deficits in dyscalculia. In D. B. Berch, & M. M. M. Mazzocco (Eds.), *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabililties* (pp. 65−81). Baltimore, MD: Paul H. Brookes Publishing Co.

Cowan, N. (1995). *Attention and memory: An integrated framework.*New York: Oxford University Press Oxford Psychology Series. Ref Type: Serial (Book, Monograph).

De Rammelaere, S., Stuyven, E., & Vandierendonck, A. (1999). The contribution of working memory resources in the verification of simple mental arithmetic sums. *Psychological Research/Psychologische Foschung*, *62*, 72−77.

De Rammelaere, S., Stuyven, E., & Vandierendonck, A. (2001). Verifying simple arithmetic sums and products: Are the phonological loop and the central executive involved? *Memory & Cognition*, *29*, 267−273.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, *20*, 487−506.

Furst, A. J., & Hitch, G. J. (2000). Separate roles for executive and phonological components of working memory in mental arithmetic. *Memory & Cognition*, *28*, 774−782.

Gathercole, S. E., & Pickering, S. J. (2000). 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.

Geary, D. C. (1990). A componential analysis of an early learning deficit in mathematics. *Journal of Experimental Child Psychology*, *49*, 363−383.

Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin*, *114*, 345−362.

Geary, D. C., Bailey, D. H., Littlefield, A., Wood, P., Hoard, M. K., & Nugent, L. (2009). First-grade predictors of mathematical learning disabilities: A latent class trajectory analysis. *Cognitive Development*, *24*, 411−429.

Geary, D. C., & Hoard, M. K. (2001). Numerical and arithmetical deficits in learning-disabled pupils: Relation to dyscalculia and dyslexia. *Aphasiology*, *15*(7), 635−647.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development*, *78*, 1343−1359.

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature*, *447*, 589−592.

Ginsburg, H. P. (1997). Mathematics learning disabilities: A view from developmental psychology. *Journal of Learning Disabilities*, *30*, 20−33.

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*, 665−668.

Hecht, S. A. (2002). Counting on working memory in simple arithmetic when counting is used for problem solving. *Memory & Cognition*, *30*, 447−455.

Hecht, S. A., Torgensen, J. K., Wagner, R. K., & Rashotte, C. A. (2001). The relations between phonological processing abilities and emerging individual differences in mathematical computation skills: A longitudinal study from second to fifth grades. *Journal of Experimental Pupil Psychology*, *79*, 192−227.

Hitch, G. J. (1978). The role of short-term working memory in mental arithmetic. *Cognitive Psychology*, *10*, 302−323.

Hitch, G. J., & McCauley, E. (1991). Working memory in pupils with specific arithmetical difficulties. *British Journal of Psychology*, *82*, 375−386.

Iuculano, T., Tang, J., Hall, C. W. B., & Butterworth, B. (2008). Core information processing deficits in developmental dyscalculia and low numeracy. *Developmental Science*, *11*(5), 669−680.

Jordan, N., & Montani, T. (1997). Cognitive arithmetic and problem solving: A comparison of pupils with specific and general mathematics difficulties. *Journal of Learning Disabilities*, *30*(6), 624−634.

Keselman, H. J., & Rogan, J. C. (1980). Repeated measures F tests and psychophysiological research: Controlling the number of false positives. *Psychophysiology*, *17*, 499−503.

Kirby, J. R., & Becker, L. D. (1988). Cognitive components of learning problems in arithmetic. *Remedial and Special Education*, *9*, 7−16.

Knops, A., Nuerk, H. -C., Fimm, B., Vohn, R., & Willmes, K. (2006). A special role for numbers in working memory? An fMRI study. *NeuroImage*, *29*, 1−14.

Koontz, K. L., & Berch, D. B. (1996). Identifying simple numerical stimuli: Processing inefficiencies exhibited by arithmetic learning disabled children. *Mathematical Cognition*, *2*, 1−23.

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8–9 year old students. *Cognition*, *93*, 99−125.

Lemaire, P., Abdi, H., & Fayol, M. (1996). The role of working memory resources in simple cognitive arithmetic. *European Journal of Cognitive Psychology*, *8*, 73−103.

Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory and Cognition*, *22*, 395−410.

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.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49−100.

Moro, R. (2008). Control and Maintenance Processes in Working Memory: Neuropsychological Investigations. PhD thesis University College London.

Noël, M. P., Seron, X., & Trovarelli, F. (2004). Working memory as a predictor of addition skills and addition strategies in children. *Current Psychology of Cognition*, *22*, 3−25.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*, 255−287.

Paivio, A., Walsh, M., & Bons, T. (1994). Concreteness effects on memory: When and why. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *20*, 1196−1204.

Palladino, P., Cornoldi, C., De Beni, R., & Pazzaglia, F. (2001). Working memory and updating processes in reading comprehension. *Memory and Cognition*, *29*, 344−354.

Passolunghi, M. C., & Siegel, L. S. (2001). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology*, *88*(4), 348−367.

Pickering, S., & Gathercole, S. (2001). *Working memory test battery for children (WMTB-C) manual.* London: Psychological Corporation Ltd.

Russell, R. L., & Ginsburg, H. P. (1984). Cognitive analysis of pupils' mathematical difficulties. *Cognition and Instruction*, *1*, 217−244.

Semenza, C., Miceli, L., & Girelli, L. (1997). A deficit for arithmetical procedures: Lack of knowledge or lack of monitoring? *Cortex*, *33*(3), 483−498.

Seyler, D. J., Kirk, E. P., & Ashcraft, M. H. (2003). Elementary subtraction. *Journal of Experimental Psychology. Learning Memory and Cognition*, *29*, 1339−1352.

Shalev, R. S., & Gross-Tsur, V. (2001). Developmental dyscalculia. Review article. *Pediatric Neurology*, *24*, 337−342.

Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development*, *60*, 973−980.

Swanson, H. L., & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in pupils with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Pupils Psychology*, *79*, 294−321.

Wechsler, D. (1996). *Manual for the Wechsler Intelligence Scale for Children* (3rd edn). London: The Psychological Corporation: Harcourt Brace & Company.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale* (3rd edition). San Antonio: The Psychological Corporation.