The Relationship of Domain-General Serial Order Memory and Reading Ability in School Children with and without Dyslexia

Wibke Maria Hachmann, Nathan Cashdollar,

Francesca Postiglione, & Remo Job

Department of Psychology and Cognitive Science, University of Trento, Italy

Center of Mind and Brain Science, University of Trento, Italy

Author Note:

Wibke Maria Hachmann, Nathan Cashdollar, Francesca Postiglione, and Remo Job, Department of Psychology and Cognitive Science, University of Trento.

Wibke M. Hachmann is now affiliated to the Department of Pedagogical Psychology, University of Halle-Wittenberg, Nathan Cashdollar is additionally affiliated to the Center of Mind and Brain Science, University of Trento, Italy, and Francesca Postiglione is additionally affiliated to the non-profit research foundation Marica De Vincenzi ONLUS, Rovereto, TN, Italy. The non-profit Fondazione Caritro, TN, Italy, kindly provided funding for this research.

Correspondence concerning this paper should be directed to: Wibke M. Hachmann, Pedagogical Psychology, Martin Luther University Halle-Wittenberg, Germany. E-mail: wibkehachmann@gmx.net, Mobile: +49 (0)1577 57 47 927.

Acknowledgements:

This research was supported by grant IPALD (Indicatori Precoci dell'Abilità di Lettura e della Dislessia; *Early indicators of reading ability and developmental dyslexia*) from the

Fondazione Marica De Vincezi ONLUS through Fondazione Cassa di Risparmio Trento e Rovereto, Italy. The first author received a scholarship from the state of Italy. The funding sources were not involved in any research related steps of this work. We are grateful to all the participants for their dedication, to the schools of Rovereto Nord and Volano, Trentino, Italy, to the teachers and parents who made the cooperation possible, and to the therapists of A.P.S.P. - Azienda per i Servizi alla Persona Beato de Tschiderer, Trento, in person to Dott.ssa Emanuela Paris, and Associazione Italiana Dislessia (AID), Sezione di Trento.

The Relationship of Domain-General Serial Order Memory and Reading Ability in School Children with and without Dyslexia

The development of reading and writing abilities in children not only requires the maturation of numerous complex linguistic skills, but also other underlying cognitive abilities that are essentially non-linguistic in nature. Children who display difficulties in reading performance and children with developmental dyslexia (DD) have demonstrated impairments in at least one, but very often several, basic underlying cognitive functions in non-verbal domains (Pennington, 2006), in addition to their deficits related to reading and writing abilities. In recent years, it has been suggested that one's ability to encode and briefly retain the serial order of information from the external environment may in fact underlie reading development (Leclercq & Majerus, 2010; Majerus & Cowan, 2016; Majerus, Poncelet, Elsen, & van der Linden, 2006). Short-term memory (STM) for the serial order of information has been linked to the ability of paired-associate learning when acquiring new words (Majerus, Poncelet, Elsen, et al., 2006), and this ability can predict the degree of vocabulary development over time (Leclercq & Majerus, 2010). Also, deficits in the capacity to retain the correct sequence of information in both linguistic and non-linguistic domains have been found in adults as well as in children with DD (Bogaerts, Szmalec, Hachmann, Page, & Duyck, 2015; Cowan et al., 2017; Hachmann et al., 2014; Martinez Perez, Majerus, Mahot, & Poncelet, 2012; Poncelet, Schyns, & Majerus, 2003; Szmalec, Loncke, Page, & Duyck, 2011). The short-term recognition of the serial order of object images and doodle drawings (Hachmann et al., 2014) and the reconstruction of a series of digits (Martinez Perez, Majerus, & Poncelet, 2013) was shown to be poorer in adults with dyslexia than in control participants. Also in children with dyslexia, the short-term reconstruction of a series of animal names

(Martinez Perez, Majerus, Mahot, et al., 2012) and digits (Nithart et al., 2009) as well as running spans for digits, locations, and marginally also shapes (Cowan et al., 2017) was found to be impaired compared to controls. Not only short-term retention, but also the transfer of serial order information to long-term memory was shown to be slower in adults (Bogaerts, Szmalec, Hachmann, Page, & Duyck, 2015) and children with dyslexia (Bogaerts, Szmalec, De Maeyer, Page, & Duyck, 2016) than in controls.

This all is supportive evidence for a substantial relation between serial order memory, reading development, and dyslexia, although “the corpus of research conducted on serial ordering deficits in short-term memory within children with dyslexia is surprisingly small” (Cowan et al., 2017, p. 230). One primary issue that still remains unclear is the time during development that reflects the mechanism with which the underlying serial order function contributes most to reading development. According to recent findings and theories about reading development, the moment when learners have to maintain a sequence of letters to convert them one by one into a phoneme to finally decode a word should be the most vulnerable for impaired serial order STM skills (Share, 2004; Snowling & Hulme, 2005; Ziegler, Bertrand, Leté, & Grainger, 2014; Zoccolotti, De Luca, Di Filippo, Judica, & Martelli, 2009). Therefore, the proposed relationship between serial order STM and reading should be most evident in developmental trajectories during the first years of schooling, specifically when children already know the letters and train grapheme-phoneme-conversion to bind them together in words. Starting from grade 1, the relation between serial order STM and reading skills should hence increase. After word reading is partly automatized by grade 3 or 4, the influence of serial order STM should decrease again, because by then parts of words form chunks and release STM resources (Share, 2004). In our present cross-sectional study, we address these questions using a novel experimental paradigm designed to isolate specific short-term memory abilities in children with and without DD during their first years of elementary school. This is the first study to investigate the relationship of STM performance to reading skills across grades 1-4.

Item versus Order Processing in Language

One intriguing property concerning the nature of verbal material in general is it's innate seriality (Acheson & MacDonald, 2009). An intrinsic requirement for cognitive processing in general (Houghton & Hartley, 1995) and for language in particular, is the underlying ability to encode, retain, and subsequently generate the serial order of sequences (Gervain, Nespor, Mazuka, Horie, & Mehler, 2008). The capacity to retain ordered sequences in memory makes it possible to integrate single items into their specific spatial and temporal context (Ahissar, Lubin, Putter-Katz, & Banai, 2006; Howard, Howard, Japikse, & Eden, 2006; Schuck, Gaschler, & Frensch, 2012).

Short-term memory (STM) encoding and retention has traditionally been conceptualized as a domain-specific process (i.e. independent networks supporting auditory or visual stimuli separately) that is contingent upon the particular memory content in question (Baddeley, 2003). Byrne (1981) discussed the relationship between STM, linguistic access, and reading skills accordingly as a critical confound of a primarily language-related triangle that needs disentangling. In a similar view, Cohen, Netley, and Clarke (1984) found a recency deficit in reading-disabled children compared to age matched controls using an analysis on segments of a digit span task, i.e. verbal material without a distinction of item and order processing.

In recent theories and models, the retention of the seriality was re-conceptualized as a domain-general function of STM that serves to bind specific items in a contextual sequence or temporal pattern, forming a new set of information (Brown, Preese, & Hulme, 2000; Burgess & Hitch, 1999; Gupta, Lipinski, & Aktunc, 2005; Majerus et al., 2010; Page & Norris, 2009). The order function is separated from item content and forms an integral part of short-term memory, while item memory represents short-term activation of long-term memory content. The combination of order and item memory together finally supports language skills. With these models, it becomes feasible to address Byrne's triangle again, because serial order STM represents a basic function that serves the linguistic and non-linguistic domains, while only the actual item content is directly related to the linguistic domain. The general function of serial order in STM, therefore, is a likely candidate for an early predictor of language development in children while avoiding the confounding of linguistic item memory, reading experience, and STM capacity discussed by Byrne.

Storing the serial order of verbal materials (i.e. phonemes, letters, morphemes, words) has been proposed as an essential skill necessary for the development of linguistic and grammatical abilities (Acheson & MacDonald, 2009; Banai & Ahissar, 2010; Majerus & Cowan, 2016; Page & Norris, 2009; Tomblin, Mainela-Arnold, & Zhang, 2007). Word order is a general property of language which is thought to develop before infants can begin to substantially build their lexicon (Gervain et al., 2008). For instance, Majerus and his team repeatedly showed that the acquisition of new vocabulary was linked to the ability of serial order reproduction (Majerus & Boukebza, 2013; Majerus, Poncelet, Elsen, et al., 2006; Majerus, Poncelet, Greffe, & Van Der Linden, 2006; Majerus, Poncelet, Van Der Linden, & Weekes, 2008; Majerus, Van der Linden, Mulder, Meulemans, & Peters, 2004). Specifically, Leclercq and Majerus (2010) demonstrated that performance on serial order STM tasks was related to the speed and learning slope of new word acquisition, while memory for item content better predicted the individual knowledge of linguistic units (i.e. phonology of a language). Accordingly, impairments in the retention and reproduction of sequences have been suggested to be linked to Specific Language Impairment (SLI) (Majerus et al., 2009; Nithart et al., 2009), but also to reading development at the onset of instruction (Martinez Perez, Majerus, & Poncelet, 2012).

Serial Order Retention and Reading Development

The ability to process the serial order of information has proven to be an important predictor not only for language acquisition but also for reading development. For instance, a longitudinal study by Martinez Perez, Majerus, and Poncelet (2012) showed that STM skills for sequential order in French speaking kindergarten children predicted later print-to-sound decoding abilities in first graders better than verbal item processing abilities. These results support the hypothesis that the capacity to maintain serial order information in STM can contribute or interact with the ability of learning how to read. Furthermore, this is in line with similar findings demonstrating a pronounced impairment for the recognition and reproduction of sequences in adults and children with dyslexia (Bogaerts, Szmalec, Hachmann, Page, & Duyck, 2015; Bogaerts, Szmalec, Hachmann, Page, Woumans, et al., 2015; Cowan et al., 2017; Hachmann et al., 2014; Martinez Perez, Majerus, Mahot, et al., 2012; Martinez Perez et al., 2013; Szmalec et al., 2011). Also besides this evidence from studies dedicated primarily to reading, children with DD have also been reported to experience difficulties related to other domains of serial order processing, such as time perception or rhythmic pattern reproduction (Leong & Goswami, 2014; Pagliarini et al., 2015). Indeed, early indicators of DD are primarily characterized by children’s restricted mnestic resources (DSM-IV) - i.e. low capacity in short and long term memory - a problem which profoundly impacts a variety of learning processes that are not specific to the domain of language (Smith-Spark & Fisk, 2007). Overall, these findings strongly suggest that children with DD might possess a basic deficit in the domain-general ability of sequential information processing in STM which, during the course of development, may manifest in increasing difficulty to automatize the reading process.

During the early stages of written language development, mnestic impairments should be mostly evident in tasks tapping into sublexical grapheme-phoneme conversion (GPC) skills, because arguably these require considerable serial order processing. In shallow orthographies, this is a learning stage in which pronunciations are derived directly from print, before syllabic, morphological, and lexical information start to support word recognition (Share, 2004). In the very transparent orthography of Italian, children master grapheme-phoneme decoding primarily during the second school year, whereas word reading is largely automatized by grade 4 (Zoccolotti et al., 2009). In between, learners have to train the initial GPC skills in speed and accuracy to a sufficient degree to arrive at stable word recognition and orthographic word production (Share, 2004). This window between second and fourth grade, therefore, should be critical for the relationship of serial order STM and reading development and a deficit thereof should be reflected in children with dyslexia.

Rationale

In accordance with this rationale, we assessed serial order and item processing abilities during early reading development in children using a balanced order and item task design in two experiments: Experiment I aimed to a) determine if serial order STM performance is related to reading development during the critical time of grades 2 and 3 in elementary school, and b) if so, whether STM abilities for serial order information interact with reading development in a domain specific (i.e. verbal STM) or a domain general manner (i.e. both verbal and nonverbal STM tasks). In Experiment 2, we sought to elucidate whether these serial order STM tasks could be used to further characterize the reading impairments in individuals with developmental dyslexia (DD) of the same age group.

According to the computational and empirical models of STM that conceptualize serial order as a domain-general function (Majerus et al., 2008; Page & Norris, 2009), children should be able to benefit from verbal and nonverbal serial order skills during the development of grapheme-phoneme decoding. In line with this argument, serial order processing should be related to reading skills primarily during the training stage of GPC skills, that is, during grades 2 and 3 of elementary school, fading away already by grade 4. Due to the very transparent nature of the Italian orthography, however, phonetic reading strategies are explicitly taught as soon as in first grade. In Experiment I we therefore investigated the relation between serial order STM and reading skill cross-sectionally in a full range of readers cohort (except for children with DD) spanning from grades 1 through 4. In Experiment II we assessed possible differences in this STM-reading relation between normal and impaired reading development in a matched sample of children with DD from the same province, taking DD as an indicator of complex delayed or deviating reading development.

Throughout this study, all STM measures were made part of one controlled design to allow comparable performance between item and order tasks and between verbal and nonverbal materials. This produced a balanced 2 x 2 design of item and order STM tasks with verbal and nonverbal material using object pictures and doodle drawings. The tasks were created such that they require serial order or item functions as exclusively as possible by primarily manipulating task instructions, but still using exactly the same experimental procedure for each condition. In using this exact same procedure with counterbalanced materials between tasks, this design is novel in the field of reading research. In recognition tasks for serial order, participants were asked whether the order of two pictures or doodle drawings was the same as in the list they had just seen. In item tasks, participants were asked whether the two pictures or doodle drawings had both been present in the list or not, irrespective of list position.

We carefully made sure that none of the materials cognitively accessed item information that might be impaired due to a lack of reading development, such as phonemes or rhymes. This was important to investigate the theoretical question of domain-generality of the serial order function adopted here – a question that is far from common agreement but that can decide between model classes and therefore crucially benefits from replications.

Experiment I: serial order STM and reading development

We used word and pseudoword reading tests and the 2x2 design of order and item STM tasks with verbal and nonverbal materials plus a control test for cognitive processing. The order and item tasks (Hsieh, Ekstrom, & Ranganath, 2011) were adapted to maximize the distinction of order versus item processing crossed with verbal and nonverbal materials, using one procedure in all four conditions contrary to previous designs (Hachmann et al., 2014; see also Majerus & Cowan, 2016 for a discussion; Majerus, Poncelet, Elsen, et al., 2006; Majerus et al., 2008). The design of this experiment manipulated the factors Task (items/order) and Domain (verbal/nonverbal). Counterbalancing the order of all four conditions, the tasks were presented to 113 children of the 1st to 4th grade of two different elementary schools in the same province.

If STM for order is related to reading performance, this should be reflected through all four grades by covariance between reading performance and correct responses in the STM order tasks. Specifically, if serial order STM is particularly important during the phase of grapheme-phoneme decoding, the relationship should be strongest in grades 2 and 3. According to the hypothesis of a domain-general serial order function in STM, this should hold true for verbal as well as nonverbal material. Importantly, the item tasks implemented here used exactly the same stimulus materials as the order tasks. But contrary to order tasks, performance on item STM tasks should not be related to reading speed or accuracy, because there is no evidence that STM for non-speeded picture recognition represents verbal or nonverbal predictor skills for reading (Landerl & Wimmer, 2008).

Method

Participants. One hundred and thirteen children from grades 1 through 4 of two elementary schools in the same Italian province participated in this experiment. For every child included in this study, parents had given written informed consent for their child’s participation. Children were included irrespective of first language, reading proficiency, or handedness. Those participants with neurological health problems (2) or those with too little vocabulary knowledge of Italian who failed to spontaneously name all of our sample pictures correctly (2) were excluded from all subsequent analyses. Another two datasets were excluded post-hoc, one because performance was below chance level (< 50% correct in either of the STM tasks) and one due to technical problems during task execution. The 107 children of the final dataset had a mean age of 8.32 years (mean (*M*) = 99.78 months, Standard deviation (*SD*) = ±11.81, range = 75 - 188) at the time of testing. Their demographic data for each school grade are presented in Table 1.

< Insert Table 1 about here >

The group sizes in grades correspond to the distribution of grades in the DD sample that is reported in Experiment II and for which this sample later served as a control group (see Experiment II, Methods). For all analyses across groups or grades we used linear mixed effects models that calculate a separate random slope for each participant and item, which represents a robust adjustment that allows comparisons of different group sizes (Bates et al., 2015; Janssen, 2012; Martín & Pérez, 2008). Each child was tested individually in a quiet room in the collaborating school. This study was approved by the ethics committee of the University of [deleted for blind review] through approval number 2012-006 for the project entitled "Studio sui meccanismi alla base dei disturbi precoci del linguaggio: una indagine su bambini in età prescolare e scolare" [Study on the basic mechanisms of early language disorders: an investigation with preschool and school children].

Material and Design. Two serial order recognition tasks and two item tasks were each furnished either with nameable pictures and words of familiar objects or with iconic nonsense drawings (see Figure 1).

< Insert Figure 1 about here >

Similar to the task for order and item memory reported in Hsieh et al. (2011), this double probe task was used to ensure equal retrieval phases and task sensitivity in all conditions. However, in a pilot version of this experimental design we noticed that order tasks were still slightly more difficult to perform than item tasks, so that the hypothesized relation between reading and serial order task performance might be attributable to general cognitive strength or overall motivation rather than to the experimental manipulation. In order to balance the level of difficulty between tasks, we therefore rendered item tasks more difficult by increasing list length (6 items vs. 4 items) and reducing presentation time (1 sec. vs. 1.5 sec.) as compared to order tasks.

Each of the four short-term memory tasks comprised a set of 24 items, summing up to 48 black and white nonsense drawings for the two nonverbal tasks and 48 colour pictures (Rossion & Pourtois, 2004) with the corresponding object names recorded into audio files for the verbal tasks. Care was taken to choose these nonsense drawings so that they had medium visual complexity and as little resemblance to existing symbols as possible. These items had been piloted and rated in a previous study [reference deleted for blind review]. Images and words were chosen on the basis of naming agreement, word frequency, and visual familiarity for this specific children sample, comprising only 2-syllable words with a mean $\log_{10}$ frequency of 2.05 (SD = ±0.65, range = 0.6 - 3.9), a naming agreement of 96% (SD = ±4 %, range = 85 % - 100 %) and an age of acquisition of 3 years and 4 months (SD = ±1 year, range in years; months = 1;8 - 5;3, see Nisi, Longoni, & Snodgrass, 2000). The names of the pictures were recorded by a female native speaker who naturally adopts the Italian standard variation of the testing region.

The sets of images and nonsense drawings were each counterbalanced between tasks, so that one set of 24 stimuli was used in the item task for one child and in the order task for another. Within conditions, materials were randomised so that each item appeared equally often in every list position. Furthermore, each item was used once as a target in every one of two target positions and all items were counterbalanced across trial lists.

Word and pseudoword reading skills were assessed with subtests 2 and 3 from the DDE-2 (Battery for the assessment of developmental dyslexia and dysorthography in Italian, second edition, Sartori, Job, & Tressoldi, 2007). The word reading test consists of 116 words printed in one column on four pages that are to be read as fast and accurately as possible. The experimenter measures the total time to read all words and registers reading errors. The procedure is exactly equal for the 51 pseudowords in the pseudoword reading task. The test provides norms from grade 2 onwards. In this study, all children from grades 2 to 4 performed the entire reading tasks. Children of grade 1 were asked to read only the first page of each test, i.e. 28 words and 16 pseudowords, to avoid fatigue and demotivation. To maintain scaling, we used raw error rates (percentage of total words read) and reading speed (syllables per second for correctly read words) of all children in all analyses (see Table 2). We are conscious of the problem that these measures originated from different test lengths for which scores might be influenced by proactive interference or stamina. However, test insensitivity at the low end is a common problem when testing low-performing populations and also holds for severe dyslexia. In this context, which had to be coordinated also with the therapy centre's test criteria, the raw scores represented the best possible solution for vertical scaling.

As a control measure for general cognitive ability, we administered the subtest *block design* of the WISC-III (Wechsler Intelligence Scale for Children, third edition in Italian, Wechsler, 1991).

Procedure. The experimental session was organized as follows: two of the four experimental tasks were carried out in the first part, followed by tests on reading abilities and the block design test. Finally the session ended with the remaining two experimental tasks.

The order of the four memory tasks was counterbalanced across participants. The standardized tests (reading words and pseudowords and the block design test) were always administered in the same order, starting with reading tests and proceeding with block design. Testing took place during February and March 2013, when children had already received half a year of school education in their grade. Each short-term memory task lasted about 7 minutes and the entire experimental session took about 45 minutes.

The memory tasks were implemented in E-Prime and all trials began with a fixation cross displayed for 500ms to direct attention to where the first stimulus was to appear to the left in the upper half of the screen. There were 18 experimental trials in each task.

In item tasks, a list of 6 items was then presented sequentially, non-cumulatively, and spaced without overlap in the upper half of the computer screen for one second per item. After list presentation, another fixation cross appeared for 500ms, this time in the screen centre. In the nonverbal item condition the fixation cross was followed by two items spaced temporally in central position for 1 second each. In the verbal item condition the two target items were presented auditorily via headphones, playing two words consecutively with the onset at 1 sec. each. Children had been instructed to decide by button press whether or not both items had been present in the list before or not. Thus, if only one or none of two items had been present, the correct answer was a *no*-response.

Order tasks followed exactly the same procedure, with the exceptions that a) lists contained 4 items instead of 6 and b) all items and targets were presented 1.5 seconds each instead of 1 second. This was done to adjust task difficulty between item and order tasks as evaluated in a pilot study with 20 children. In order tasks, children were instructed to decide whether or not the first target had indeed been displayed before the second one in the list. Target items never represented two adjacent items from the list; therefore a probe of positions

1 and 3, for example, would mean a *yes*-response while positions 4 and 1 would require a *no*-response.

The maximum possible time to answer was 10 seconds. After that, a message appeared to announce the next trial which was self-paced by the participant. A pre-test of 6 trials was done prior to the experiment to ensure the children understood the procedure and became familiar with the kind of stimulus presentation and response buttons, using the same stimulus material and providing feedback for each response. However, during the experiment proper no feedback was given.

Although the differentiation of tasks has been evaluated before (Hsieh et al., 2011), memory for serial order information could still be recruited for item task performance to some extent. There was no explicit requirement of serial order processing for the item tasks implemented here, however, item tasks are often implicitly sensitive to serial order strategies to facilitate performance (for a discussion on psychometrics see Cowan et al., 2017). For instance, it might be easier to remember an item if the list position of the item is stored as additional memory cue to its identity information. To control for the influence of order memory on item task performance, the order of those targets in item tasks that constituted a *yes*-trial (all trials in which both target items had been present in the encoding list before) was balanced between chronologically ordered target presentation and presentation in opposite order, and this chronological order per trial was tracked for later analysis (see the Results section on the 'contribution of serial order processing to item task performance'). Also, the targets that originated from the encoding lists were counterbalanced between these list positions. In all experimental conditions, accuracy was recorded as the dependent measure.

Data Analysis. Data was analyzed in R, open source software for statistical analysis and mathematical computing (R version 3.3.1, R Core Team, 2016). For descriptive statistics, we calculated the percentage of correct responses by participant for each of the four experimental conditions. For the same conditions, we derived mean reaction times for trials with correct responses only.

First, we present an overview of reading performance across grades using linear regression analyses. The variable Reading later formed the continuous covariate for the analysis on STM performance. STM performance was analyzed with analyses of variance (ANOVAs) on accuracy data based on mixed effects regression models, applying the assumption of a Binomial distribution for raw accuracy data. This procedure guarantees adequate estimations of raw data in various group sizes while still ensuring conservative error statistics. For each participant, the model estimates fixed and random effects in the function *glmer* (R-package 'lme4', Bates et al., 2015). The fixed factors Task and Domain were coded as effect contrasts (for example: order tasks = 1, item tasks = -1). Similar to ANOVAs, the regression intercept therefore represents the overall mean response accuracy to which each variable coefficient is contrasted, producing orthogonal effects and interactions. The factor Grade was coded with difference contrasts (also called *reverse Helmert coding*), contrasting each higher grade to all previous grades. This is an appropriate method for ordered categorical variables that is especially useful in a developmental context. For each step in development, it allows comparisons of change in achievement to previous steps. Participant number and trial list were entered as random factors to control for specific item combinations or individual participant effects and to separate out group size differences (Martín & Pérez, 2008; Stein & Stanford, 2008). With the function *Anova* (R-package 'car', Fox et al., 2016) we then used the effects of the linear model for a type III model comparison. These ANOVA results constitute the basis of our report. Because they provide even more conservative results than the mixed models themselves, other *glmer* effects were discarded. Post-hoc details for the single contrasts within the ANOVA were extracted using the function *glht* (R-package 'multcomp', Hothorn et al., 2016) on the linear mixed effects regression models with cell mean coding. For details of interaction contrasts involving continuous variables (reading covariates) we used the function *lht* (R-package 'car', Fox et al., 2016).

To analyze STM performance by Grade, we conducted an omnibus ANOVA on linear mixed effects models of accuracy on the fixed factors Task (order/item), Domain (verbal/nonverbal), and Grade (1-4) and their interactions, with random slopes for participant and trial list. The relation of STM performance to reading abilities across Grade was then investigated in further models that additionally included one of the four z-transformed continuous covariates reading speed or accuracy for word or pseudoword reading.

Results

Reading Skills across Grade. Raw reading times were divided by the total number of syllables read to provide a reading speed measure of syllables per second per child separately for word and pseudoword reading. Reading accuracy represents the proportion of errors per total words or pseudowords for each participant. Performance for each measure is presented in Table 2, showing syllables per second, error percentages, and test norms for each grade where applicable.

< Insert Table 2 about here >

Overall, children in this sample tended to read very accurately with a slight cost of reading speed compared to the larger norm sample from the diagnostic test basis, but there were no significant differences and therefore our experimental sample can be considered a representative group of normal readers. To investigate the developmental pathway of reading and short-term memory functions, we used raw syllables per second and error percentages in all further analyses. These served as dependent variable when reporting reading outcome across grades, and as continuous covariates when evaluating the relations between serial order STM and reading.

To inquire how reading performance developed across grades, we ran linear regressions for error percentages and reading speed (syllables per second) separately for word and pseudoword reading on the factor Grade. Grade was coded with deviance contrasts for ordered categorical variables, allowing that each next larger grade level is compared to all previous younger grades. As expected using this approach, our results demonstrate a relative increase in skills across Grade, see Table 2.

The largest development in reading times and accuracy was evident between grades 2 and 3 (accuracy: $\beta = 1.47$, $SE = .33$, $t = 4.48$, $p < .001$; reading speed: $\beta = .31$, $SE = .02$, $t = 12.63$, $p < .001$). In word reading, children became faster at grade 2 with respect to grade 1, $\beta = .22$, $SE = .05$, $t = 4.59$, $p < .001$. At grade 3, children read words even faster, $\beta = .41$, $SE = .03$, $t = 12.42$, $p < .001$ and also more accurately, $\beta = 1.18$, $SE = .22$, $t = 5.47$, $p < .001$, which is maintained by children from grade 4 (accuracy: $\beta = .33$, $SE = .11$, $t = 2.95$, $p < .01$; syllables per second: $\beta = .19$, $SE = .02$, $t = 11.56$, $p < .001$).

For pseudoword reading, children in grade 2 read faster than their fellows from grade 1, $\beta = .15$, $SE = .03$, $t = 5.06$, $p < .001$, but this speed was achieved at the cost of making significantly more errors, $\beta = 3.23$, $SE = .71$, $t = -4.55$, $p < .001$. This trade-off was resolved in grade 3, where children read pseudowords faster, $\beta = .22$, $SE = .02$, $t = 10.95$, $p < .001$, and also more accurately, $\beta = 1.76$, $SE = .48$, $t = 3.66$, $p < .001$, than in previous grades, in line with the findings from word reading. In grade 4, pseudoword reading speed was kept up with respect to previous grades, $\beta = .10$, $SE = .01$, $t = 10.53$, $p < .001$, while accuracy did not show further differences to all previous grades, $\beta = .31$, $SE = .24$, $t = -1.28$, $p = .203$.

Looking more closely into why accuracy performance stagnated in fourth grade, we discovered that fifteen out of 39 fourth-grade children spoke another language beside Italian on a regular basis. This relatively large number of varying types of bilinguals might have contributed to weaker reading performance. Indeed, when taking into account also the twostepped effect-coded factor Bilingualism, in 4th grade the linear regression revealed a negative interaction of Bilingualism with pseudoword reading accuracy, $\beta = 2.772$, $SE = .634$, $t = -4.375$, $p < .001$, and speed, $\beta = -.085$, $SE = .025$, $t = -3.352$, $p < .001$, and with reading speed for words, $\beta = -.126$, $SE = .042$, $t = -3.022$, p < .01. Bilingual children in fourth grade in our sample made more pseudoword reading errors and read more slowly than monolinguals. This fact might occlude the findings we were investigating, i.e. that reading and serial order STM processing interact strongest during grades 2 and 3, by overestimating the serial order STM capacities needed in grade 4 in bilingual children. The time period in which serial order memory and reading are most dependent might differ between orthographic systems, because the point at which GPC decoding skills are most important should differ according to language specific constraints in the time-course of orthography acquisition (Ziegler & Goswami, 2005). We therefore ran model comparisons with and without the factor Bilingualism for the mixed effect models of memory task performance and reading that are reported below. There were no significant differences between models with or without the factor Bilingualism (all $p > .1$). Also, including only monolingual participants in the analysis on STM tasks resulted in the same significant effects. We therefore kept all data to represent a common modern sample of elementary school children. Results of models controlling for Bilingualism are provided in supplementary materials.

STM Performance across Grade. The mean number of correct responses for each grade and condition are presented in Table 3. Because the models are run on raw accuracy data, Table 3 presents the numbers of correct responses as the descriptive format of raw data. To depict inference statistics in Figure 2 below, we then used the more common percentage of correct responses that were back-transformed from logits of the Binomial model. Tables of all model results can be found in the appendix.

< Insert Table 3 about here >

To analyze STM performance in relation to school grade, we performed analyses of variance (ANOVAs) on linear mixed effect models of accuracy on the fixed factors Task (order/item), Domain (verbal/nonverbal), and Grade (1-4) with random intercepts for participant (n = 107) and trial list (4 x 24 lists). Means and standard deviations in post-hoc contrasts are back-transformed from model logits that are accounting for random effects and all other factors. Raw accuracy data are presented in Table 3.

There were main effects of Task, $\chi^2(1) = 123.7$, $p < .001$, of Domain, $\chi^2(1) = 67.34$, $p < .001$, and of Grade, $\chi^2(3) = 19.08$, $p < .001$. As planned by our experimental design, serial order tasks (*M* = 80.74 %, *SE* = 2.18) were easier to perform than item tasks (*M* = 67.96 %, *SE* = 2.8). Also, performance on verbal material (*M* = 79.05 %, *SE* = 2.26) was better than on nonverbal material (*M* = 69.65 %, *SE* = 2.72), as represented in the main effect for Domain. Third, the main effect of Grade showed that fourth graders (*M* = 78.1 %, *SE* = 1.82) performed better overall than children from first (*M* = 71.33 %, *SE* = 3.16, $t = 3.35$, $p < .001$) and second grade (*M* = 72.33 %, *SE* = 2.08, $t = 3.81$, $p < .001$), and statistically equally well as third graders (*M* = 75.64 %, *SE* = 2.91, $t = 0.93$, $p = .35$). This effect, however, was qualified by an interaction of Task and Grade, $\chi^2(3) = 21.22$, $p < .001$. The capacity to remember serial order information changed with school years (Δ order task performance from grade 1-4: 10%), while the ability to recognize the specific items did not change significantly from grade 1 to 4 (Δ item task performance from grade 1-4: 2 %). Figure 2 depicts the relationship of school grade and STM performance in all four STM tasks.

< Insert Figure 2 about here >

The interaction showed that in first grade, children performed only a little better in order than in item tasks, Δ order-item = 6.43 %. In second grade the difference between tasks, Δ = 11.32 %, was marginally larger than in grade 1, $t = 1.73$, $p = .08$. Children of grade 3 showed the largest Task effect, Δ = 18.29 %, which was significantly larger than that in grade 2: $t =$

2.55, $p < .05$. This level of order over item task performance was similar in fourth grade, Δ = 15.1 %, but not significantly different from grade 3, $t < .6$. Although all grades showed order above item task performance, the largest gain of this effect relative to younger grades was achieved by grade 3. This corresponds to the largest increase in reading skills, which also took place between grades 2 and 3, see Table 2. Therefore, we next looked into how the development of serial order performance and of reading skills interacted across grade levels.

Relation of Reading and STM Performance across Grade. The four measures of Word and Pseudoword reading Speed and Accuracy were z-transformed and added to the omnibus model separately. Figures 3a and 3b depict bimodal line plots respectively of word reading accuracy and speed by Grade, separately accounting for order and item task performance. Reading performance for pseudowords by grade is displayed in Figures 4a and 4b.

< Insert Figures 3a and 3b about here >

When adding the covariate Word Reading Accuracy, beside the main effects of Task, $\chi^2(1) = 66.67$, $p < .001$, Domain, $\chi^2(1) = 47.1$, $p < .001$, and Grade, $\chi^2(3) = 13.16$, $p < .01$, and the interaction of Task and Grade, $\chi^2(3) = 13.68$, $p < .01$, there was a marginal interaction of Task and Word Reading Accuracy, $\chi^2(1) = 3.81$, $p = .051$, and a three-way interaction of Task, Grade and Word Reading Accuracy, $\chi^2(3) = 7.86$, $p < .05$. The three-way interaction specified that the largest gain in order task performance happened at different grades for upper and lower parts of the reading continuum. Overall, children who made fewer word reading errors also performed better in order tasks (Δ of order - item in low reader spectrum = 8.38 %; in high reader spectrum = 14.4 %). For readers of the higher accuracy spectrum this was most pronounced in grade 3, while at the lower end of the reading distribution this difference was strongest by grade 4. As early as in first grade, the more correctly children read words, the more they already showed an advantage of order over item tasks, Δ = 10.95

%, $\chi^2(1) = 8.28$, $p < .01$. In second grade, all readers showed an advantage of order over item tasks, Δ = 10.99 %, $\chi^2(1) = 37.57$, $p < .001$. The more correctly children of grade 3 read words, the stronger they displayed an advantage of order over item tasks, Δ = 23.23 %, $\chi^2(1) = 30.3$, $p < .001$, leading the field in the overall improvement in grade 3. A similar increase in order task advantage was evident also for readers of the lower accuracy spectrum in grade 4, Δ = 18.3 %, $\chi^2(1) = 32.04$, $p < .001$, which was at least as high as for all readers of their grade, Δ = 13.4 %, $\chi^2(1) = 37.86$, $p < .001$.

When instead adding Word Reading Speed to the omnibus model, there was a main effect of Task, $\chi^2(1) = 40.54$, $p < .001$, of Domain, $\chi^2(1) = 38.82$, $p < .001$, and of Word Reading Speed, $\chi^2(1) = 7.29$, $p < .01$. There further was the familiar interaction of Task and Grade, $\chi^2(3) = 9.76$, $p < .05$, and a three-way interaction of Task, Grade, and Word Reading Speed, $\chi^2(3) = 8.62$, $p < .05$. The positive main effect of Word Reading Speed denotes that faster word readers also performed higher in all STM tasks, Δ = 5.42 %, $t = 2.7$, $p < .01$. The three-way interaction showed that third grade children displayed more of an order task advantage the faster they read words, Δ = 19.86 %, $\chi^2(3) = 45.32$, $p < .001$. This yielded one significant interaction contrast for order versus item tasks in the word reading speed slope for grade 3 versus younger grades, Δ = 15.46 %, $t = 2.77$, $p < .01$. This is in line with the results of word reading accuracy, in which the more correctly children read, the more they also showed a gain in order task performance, particularly in grade 3.

Accounting for Pseudoword Reading Accuracy yielded the familiar effects of Task, $\chi^2(1) = 54.43$, $p < .001$, Domain, $\chi^2(1) = 31.7$, $p < .001$, and Grade, $\chi^2(3) = 18.81$, $p < .001$, and the interaction of Task and Grade, $\chi^2(3) = 14.59$, $p < .01$, but no effects or interactions with Pseudoword Reading Accuracy (all $p > .099$).

For Pseudoword Reading Speed, however, there was an effect of Task, $\chi^2(1) = 65.39$, $p < .001$, and Domain, $\chi^2(1) = 33.41$, $p < .001$, the interaction of Task and Grade, $\chi^2(3) = 9.02$, p

< .05, and a main effect of Pseudoword Reading Speed, $\chi^2(1) = 7.65$, $p < .01$. As in Word Reading Speed, the faster children read pseudowords, the higher they performed in all STM tasks, $\Delta = 5.03$ %, $t = 2.77$, $p < .01$. The relation of STM performance by Grade according to Pseudoword Reading Accuracy and Speed is presented in Figures 4a and 4b.

< Insert Figures 4a and 4b about here >

Contribution of Serial Order Processing to Item Task Performance. Information of serial order was also present in our item tasks, but depending on the individual's encoding strategy, this may not necessarily have influenced performance. For instance, in item tasks, a trial was judged correct if both targets originated from the encoding list. Hence, *yes*-trials incorporated a relative order position from the encoding list and targets were either shown in correct or opposite order, just as in serial order tasks. The forward versus backward information may possibly have supported the recognition of familiar items in addition to identity information, whereas missing order information might have boosted an item's recognition as unfamiliar. Since this analysis includes subsets of data by participant, the statistical power is lower. Therefore, we propose these analyses as additional tentative evidence, not as final conclusive results.

Data of item tasks were split to create the new factor Chronology. If the first target in *yes*-trials originated from a position before the second target, the trial was considered a *forward* trial, and the reverse a *backward* trial. The other trials (*no*-trials) were split into three categories of equal size: 1) only the first target from the list, called *t1* trials, 2) only the second target came from the list, called *t2* trials, and 3) *none* of the targets came from the list. These five exclusive steps of chronological information constituted the sum coded factor Chronology. Table 4 provides percentages of correct responses in item tasks by Domain and Chronology.

< Insert Table 4 about here >

The ANOVA of accuracy on Chronology and Domain in item tasks revealed main effects of Chronology $\chi^2(1) = 37.03$, $p < .001$, and of Domain, $\chi^2(1) = 25.07$, $p < .001$, which were qualified by their interaction, $\chi^2(1) = 15.99$, $p < .01$ (see table of model results in the appendix). Figure 5 depicts back transformed mean percentages of correct responses by chronological target presentation for the verbal and the nonverbal item task.

< Insert Figure 5 about here >

In line with the results reported above, the main effect of Domain confirmed within item tasks that verbal material (M = 72.13 %, SE = 2.61) was easier to process than nonverbal material (M = 64.09 %, SE = 2.89). Contrasts on the main effect for Chronology first showed that t1 trials elicited lower performance compared to backward, t2 and none trials (all $t > 2$, all $p < .05$). This effect addresses false alarm errors. If the first target was familiar, it might have more often lured children into guessing the second target was from the same list too. Second, trials with only novel targets (*none*) showed significantly higher correct rejections than all other trials, even when considering the contrast to only the other *no*-trials (all $t > 3.4$, all $p < .001$). This confirms that items were more easily judged as new when they were not associated with any serial order information from the probe.

The interaction between Chronology and Domain, interestingly, showed differential effects for forward and backward trial performance in verbal and nonverbal material. Post-hoc contrasts showed that children responded to forward trials more accurately with verbal than with nonverbal material, 21.38 %, $t = 5.86$, $p < .001$. In the verbal item task, forward trials (M = 75.16 %, SE = 2.14) showed marginally better performance than backward trials (M = 68.94 %, SE = 2.6, $t = 1.86$, $p = .06$). In the nonverbal item task, on the contrary, backward trials (M = 65.37 %, SE = 2.48) were significantly more accurate than forward trials (M = 53.78 %, SE = 2.94), $t = 3.06$, $p < .001$. This opposite pattern yielded a significant interaction of forward versus backward trials with Domain, Δ = 17.8 %, $t = 3.45$, $p < .001$.

For verbal material, forward target order thus seemed to have a supportive function, while item recognition for nonverbal material benefitted more from recency, when items that were shown last appeared first again as targets. There was a significant difference also for latencies in forward minus backward trials showing that forward trials were responded to faster with verbal (Δ = -203.61 ms) and slower with nonverbal material (Δ = +37.03 ms) $t = 2.04$, $p < .05$. This confirms that the above effect in accuracy did not merely result from a speed-accuracy trade-off.

Discussion

Our first experiment investigated the relation of reading proficiency and STM performance for serial order compared to item information in elementary school children across grades 1 to 4. Using a novel experimental design that balances the STM demands of item and sequential order retention with verbal and nonverbal material, we demonstrated a relationship of reading abilities and serial order STM performance that was particularly pronounced in second and third grade.

The overall pattern of reading errors versus reading speed is congruent with findings from other transparent orthographies, in which reading accuracy plays a larger role in early school grades. Later, accuracy reaches ceiling and reading speed becomes a more reliable measure (Aro & Wimmer, 2003; Landerl & Wimmer, 2008; Wimmer & Mayringer, 2002). First, children seemed to display a cost / benefit relation between faster and more accurate reading, which for words was evident earlier than for pseudowords (see Figures 3 and 4).

Reading development corresponded to the pattern of performance in serial order STM tasks. Also in serial order tasks, there was a performance increase between grades 2 and 3. The interactions of serial order processing and reading started already in first grade. Although in grade 1 performance on serial order tasks was lowest overall, the more accurately children could already read, the better was their serial order STM performance. By second grade, all readers had caught up on an increase of serial order STM performance relative to that of item tasks. Third graders then showed a boost in word reading speed, word reading accuracy, and serial order STM performance, which was largest in high performers. By grade 4, this boost was also visible in the lower performance spectrum. In conclusion, between grades 2 and 3, where grapheme-phoneme decoding is trained to achieve word recognition, the correspondence between serial order processing capacities and reading measures was largest.

Cross-sectional design studies should be taken with some caution concerning claims for development. We therefore interpret our findings as a strong suggestion that serial order memory is an important factor in the early development of reading proficiency, particularly during GPC decoding. But we do not interpret our data as a row of developmental steps or time trajectories. Also, reading one page of the test in grade 1 versus the whole test in grades 2-4 might have influenced performance (see the Study limitations section for more detail). However, our data fit well with the literature on reading development so we cautiously discuss them as such and emphasize that more investigation, for example with other reading measures, is necessary. As introduced, the clearly dominating evidence in reading research previews that in early school years novel words need to be decoded serially by grapheme-phoneme conversion (GPC) rules (Share, 2004; Ziegler & Goswami, 2005). This process evidently imposes high demands on STM resources (Smith-Spark & Fisk, 2007) and requires stable encoding of serial information during presentation (Binamé & Poncelet, 2016). Here we demonstrate that the ability to maintain serial order in STM emerges along early reading performance right during the school years in which GPC decoding is essential, seemingly developing along the lines of efficient reading strategies. The Italian language in particular represents a highly transparent orthography, in which children start to use also lexical-semantic support already from grade 2 onward (Zoccolotti et al., 2009). Still, the conversion of GPC rules should provide the basic building block of reading (Frost, Katz, & Bentin, 1987), and according to our data, serial order STM plays a role in word reading from grade 1 in high performers through a boost in both reading and serial order SMT performance in grade 3. According to the self-teaching hypothesis of Share (2004), learners of a shallow orthography expand the single GPC skills by chunking letters into word components, which increases decoding speed and finally leads to hierarchically organized GPC rules on the level of sub-word chunks like morphemes or syllables until whole-word recognition is possible. Our results are in accordance with the view that serial order processing should be especially important during this chunking phase between single letter decoding and whole-word recognition. This phase may represent a critical temporal window for the co-development of reading skills and serial order STM functions.

In item tasks, the particular stimulus order during the probe also demonstrated an influence on performance, again replicating the important role of sequential processing for the cognitive development of children (Houghton & Hartley, 1995). Specifically, forward trials were easier to answer in the verbal item condition and backward trials were with nonverbal material (order facilitation between verbal and nonverbal item tasks: $\Delta = 34.42\%$ accuracy, see Figure 5). These differential effects of facilitation based stimulus ordering were very stable across participants, showing that only 6 out of 107 children showed a mild opposite pattern in the verbal item task (less than 20% better performance for backward trials) and 9 of 107 children showed an opposite pattern in the nonverbal item task (20.3% forward facilitation). Therefore, seriality seems to be an important factor for cognitive facilitation in list-like recognition. Further, this facilitation seems to depend upon the particular stimulus material being learned.

Forward facilitation with verbal material could be attributed to a verbal rehearsal strategy during encoding, which may reflect sequential storage and internal rehearsal of these 6 words in a simulations similar to the concept of a phonological buffer (Baddeley, 2003). This is also in line with gradually decaying activation for each list position when building a sequential representation as proposed in the model of Page and Norris (2009). For the contrary backward facilitation effect in nonverbal materials, the most parsimonious interpretation is that of a visual-spatial recency effect (Gouldthorp, Katsipis, & Mueller, 2017; Katz & Frost, 1992; Solomon et al., 1961). In the standard modality effect, a recency effect is elicited when a stimulus sequence is not processed deeply or verbally, as impossible in our nonverbal task. Hence, the most recent items may still be most active in a visual buffer and therefore either a) be easier to recall due to decaying visual buffer activity (Engle & Mobley, 1976), or b) be easier to find back on a mental representation of the spatial location (Katz & Frost, 1992; Solomon et al., 1961). In our experiment it would follow that when one of the most recent nonsense drawings was shown first as a target, it was easier to also recognize the second target as originating from the same list. Since previous evidence suggests a particularly spatial component of visual recency (Katz & Frost, 1992; Solomon et al., 1961), children in our study may have stored the nonverbal drawings on spatial locations and therefore been able to recall the last items first again. Here, we show backward facilitation in visual processing, which differs in quality to the initial repetition of final sequences shown in the study of Simons and Chabris (1999). Therefore, it may be concluded that a purely visual procedure relies on some spatial order, which is not associated to auditory processes. A visual list can be assessed from both sides, at least. Hence, if gradually decaying activation was the mechanism by which seriality was represented also for nonverbal material, the recency effect shown here may originate from a different process than what has traditionally been considered the recency effect in auditory processing (Page and Norris, 2009). In light of this difference, we propose that the representation of visual material is more flexible than that of verbal content, in that it allows ordering and adding parts to a series - or to more complex patterns, that is - from any direction.

In summary, STM performance for serial order and reading skills did not increase linearly with each school year, but they seemed to interact in a way that supports children in developing both abilities. This was especially prevalent during the second and third school year where increased STM performance for serial order was related not only to more accurate, but also to faster reading. In comparison, reading performance did not interact with item task performance in any of these measures. To the best of our knowledge this clear difference is due to using a novel experimental design specifically targeted at balancing these factors for the first time in a way that all four tasks originate from one design, which provides strong evidence for the distinction of item and order processes. The purpose here was to use item tasks that require as little reading-related item materials as possible to not confound these functions with reading experience.

Having evaluated how order and item processing in STM interact with reading acquisition across grades in a range of normal readers, we next investigated whether this interaction differs in children with developmental dyslexia (DD). First, we aimed to determine whether indeed there is a substantial difference in order task performance also between dyslexic children and controls at elementary school age, as other studies have suggested (Bogaerts et al., 2016; Cowan et al., 2017; Martinez Perez, Majerus, Mahot, et al., 2012). Second, we wanted to elucidate if this pattern would differently characterize children with DD compared to a control group that is matched for reading age. These results should provide valuable insights into the underlying cognitive deficits in children with DD, providing further answers to the question if their reading impairments are similar to younger children with comparable reading experience due to a lack of training (i.e. a delayed cognitive development), or instead represent a specific underlying deficit that is unique to children with DD.

Experiment II: serial order STM in developmental dyslexia

The second study was conducted with a group of 16 children with DD from the same region of Italy as the children in Experiment I. All children from Experiment I served as controls, split into a reading age matched control group (RA, grade 1) and a chronological age matched control group (CA, grades 2-4). Here, we additionally adopted a specific question concerning serial order STM performance and reading development according to the results from Experiment I: Do serial order STM functions relate to reading skills in DD, even if performance on both tasks may be low? There were two possible outcomes: First, serial order memory capacity could be related to reading proficiency in children with DD similar to that of the control children from Experiment I. In this case, we would expect that serial order task performance compared to controls is impaired proportionally to the lack of reading skill, however, still being related to reading ability. The second possibility was that serial order memory is found to be independent of DD's reading ability and does not relate to reading skill in any of the four measures. Also for this purpose, it was vital to use the raw STM data and raw reading errors and syllables per second to allow comparability between groups/conditions and to avoid statistical insensitivity in low performance ranges of the norm data.

Method

Participants. In addition to the sample reported in Experiment I, 16 participants with a formal diagnosis of DD were recruited from the aforementioned schools (6 children) and from the local therapy center, the logopedic and speech therapy center [name deleted for blind review] (10 children). The therapy centre provided reading and cognitive ability measures from their recent full formal diagnosis. However, we could not retrieve reading speed measures for three out of 10 children from the therapy center. The therapists informed us that they discarded the time measures because they exceeded the norm scales by far. This assures us that poor reading accuracy was not a result of a trade-off for fast reading. One of these children additionally had not at all been able to carry out the pseudoword reading task. In our control variables for reading speed and accuracy, we adopted the lowest measure for accuracy and the longest reading times of our sample instead.

The children in the DD group had been tested by accredited therapists, applying international differential diagnostic criteria that include, among others, the core diagnostic criteria for DD: unexpectedly poor reading and spelling performance below the 10^{th} percentile or below 2 SD of the norm for their age group, normal to high IQ, and the exclusion of sensory and neurological problems. We additionally excluded those participants pre-hoc who displayed co-morbidity with other learning related disorders. With these rather strict criteria, the group can be described as of medium to severe DD.

The participants we had tested in Experiment I formed the two control groups for this sample. All 15 first graders formed the reading age matched control group (RA), whereas all the other 92 children represented the group of chronological age matched controls (CA). Demographic and test data for all groups are presented in Table 5.

< Insert Table 5 here >

For each grade, the relative number of children corresponds approximately to the ratio between control and DD groups (n of DD in grade 1-4: 2, 7, 1, 6; n of controls in grade 1-4: 15, 39, 14, 39, see Table 1). Note that grade is not a factor in this group analysis but served only to match samples relative to grade level. In this way we do not select participants, but use all collected data, which results in more robust statistical power and avoids the serious problems of post-test sample selection. The discrepancy in sample-size is controlled for by applying linear mixed effects regression analyses with a random intercept for participant before using ANOVAs, as detailed above. Children with dyslexia in our sample scored equally well on reading speed as the reading age matched control group (RA), but lower on accuracy. With this severe DD sample this was unavoidable and might in part be attributable to test sensitivity issues at the low performance end.

Material and Procedure. The procedure was identical to that of Experiment I, except that children who were tested at the therapy centre did not perform the block design test and reading measures with us, but had done them with their therapists during the course of diagnosis.

Results

STM performance by Group. The mean number of correct responses to the four STM tasks for each group is presented in Table 6.

< Insert Table 6 here >

We analysed accuracy in STM tasks in an omnibus analysis on the fixed factors Task, Domain, and Group with random effects for subject and list item. Group was dummy coded (called *treatment* coding in R) with the base being the chronological age-matched control group (CA). The ANOVA showed main effects of Task, $\chi^2(1) = 161.09$, $p < .001$, of Domain, $\chi^2(1) = 58.22$, $p < .001$, and of Group, $\chi^2(2) = 13.3$, $p < .01$, and an interaction of Task with Group, $\chi^2(2) = 29$, $p < .001$. For Task and Domain, contrasts revealed the same effects as in Experiment I, in which order tasks, $\Delta = 7.62\%$, $t = 5.86$, $p < .001$, and verbal tasks, $\Delta = 9.15\%$, $t = 6.94$, $p < .001$, were easier to perform than item and nonverbal tasks. Contrasts for the main effect of Group revealed that reading age matched controls (RA) performed lower than the CA group in all STM tasks, $\Delta = -3.98$ %, $t = -2.01$, $p < .05$, and also the dyslexia group (DD) performed lower than the CA group, $\Delta = -6.54$ %, $t = -3.29$, $p < .01$, but groups D and RA did not differ significantly, $\Delta = 2.56$ %, $p > .1$. The interaction contrasts on Group and Task revealed that both control groups showed better performance on order than on item tasks (CA: Δ = 14.03 %, $t = 12.7$, $p < .001$; RA: Δ = 6.44 %, $t = 2.16$, $p < .05$), while the dyslexia group showed equal performance on both types of tasks, Δ = 2.49 %, $p > .4$. The difference between item and order tasks was larger in group CA than in both other groups (vs. DD: Δ = -11.64 %, $t = -4.78$, $p < .001$; vs. RA: Δ = -7.58 %, $t = -3.26$, $p < .01$). However, there were no differences between groups DD and RA, neither for item vs. order tasks, Δ = -4.05%, $t = -1.02$, $p = .31$, nor within only serial order tasks, Δ = -4.59 %, $t = -1.3$, $p = .19$. Further, there were no differences between groups in item tasks, all $p > .8$. These results repeat the findings from Experiment I in that the advantage of serial order task performance was not yet present in first graders (group RA: $M = 74.56$ %, $SE = 3.05$) as it was in higher grades (group CA: $M = 82.34$ %, $SE = 1.19$). They further confirm our hypothesis that children with DD performed poorer on serial order STM tasks ($M = 69.97$ %, $SE = 3.21$) than chronological age matched controls, and at least as low on serial order tasks as first graders, but not lower than other groups on tasks tapping into identity information. Figure 6 depicts the relationships of Group and STM performance as back-transformed from model coefficients, controlling for participant and list slopes. Adding results from the z-transformed continuous covariate block design resulted in a main effect of block design, $\chi^2(2) = 9.9$, $p < .01$, and else in the same effects as reported above (see supplementary materials for results).

< Insert Figure 6 about here >

As a group, children with dyslexia thus statistically performed at the same level as first graders in STM all tasks. We next looked at individual patterns. Of 15 children in first grade, 9 children showed higher performance in order than in item tasks (Δ order-item = 15.43%, $SD = 6.53$) in line with the main effect of Task, and 6 children showed the opposite pattern (Δ item-order = 7.87%, $SD = 3.25$). In group DD, 9 of 16 children showed higher performance in order than in item tasks (Δ order-item = 10.73%, $SD = 8.38$), 2 children showed exactly equal performance in both types of tasks, and 5 children showed the opposite pattern, i.e. item above order task performance (Δ item-order = 12.22%, *SD* = 3.17). Also the variation of performance in STM tasks can thus be described as fairly equal in both groups. However, these results do not reflect the individual relation of STM performance to reading. Therefore, we next analysed the relation of STM performance to reading skills in all three groups.

Serial Order task performance and reading by Group. Adding the four reading measures to the model of STM accuracy on Task and Domain for each group separately first of all showed similar results for groups RA and CA to those reported above: in first grade (group RA), serial order task performance for verbal material related positively to reading accuracy for words, $\chi^2(1) = 4.14$, $p < .05$, and for pseudowords, $\chi^2(1) = 4.00$, $p < .05$. Also, the faster first graders read pseudowords, the better they performed in all STM tasks, $\chi^2(1) = 5.12$, $p < .05$. In all further grades (group CA), reading speed measures related to general serial order STM for both word, $\chi^2(1) = 8.63$, $p < .01$, and pseudoword reading, $\chi^2(1) = 7.19$, $p < .01$.

In group DD, first there was a main effect of Word Reading Speed, $\chi^2(1) = 9.75$, $p < .01$. This indicated that the faster children with dyslexia read words, the more accurately they performed across all STM tasks (range = 10%, *M* = 74.12%, *SD* = 5.23 in fast spectrum, *M* = 64.41%, *SD* = 6.45 in slow spectrum). Second, there was an interaction of Pseudoword Reading Speed with Task and Domain in group DD, $\chi^2(1) = 4.94$, $p < .05$. Contrasts indicated that the faster children with DD read pseudowords, the better they performed particularly in the verbal order task (range = 10%, *M* = 77.43%, *SD* = 5.01 in fast spectrum; *M* = 67.58%, *SD* = 6.46 in slower spectrum). However, when submitting all groups to an omnibus analysis with each of the four reading measures, there were no significant interaction effects between Group and reading measures (all $\chi^2(n) < 5$, all $p > .08$).

The small relation of reading measures with serial order task performance did not hold for all individuals with dyslexia, though. Two children performed above average in reading measures compared to their DD group peers (words: Δ = 18.35 % and 14.78 %, 0.32 and 0.07 syll/sec; pseudowords: Δ = 9.15 % and 4.98 %, 0.8 and 0.4 syll/sec), but below average in the nonverbal serial order task (Δ = -13.19 % and -10.41 %). Their verbal serial order performance lay just below group average (Δ = -3.13 % and -0.35 %). On the contrary, one child displayed very slow reading even compared to DD group peers (words: Δ = 11.21 %, -0.27 syll/sec; pseudowords: Δ = 0.81 %, -0.22 syll/sec), but above average performance on the verbal serial order task, Δ = 8.0 % (nonverbal serial order: Δ = -0.7 %).

Contribution of serial order information to item task performance by Group. We used the same procedure as in Experiment I to evaluate the implicit order information that might have contributed to item task performance, using the factor Group instead of Grade. Similar to the results of Experiment I, there were main effects of Domain, $\chi^2(1) = 14.56$, $p < .001$, of Chronology, $\chi^2(4) = 44.66$, $p < .001$, and an interaction of Domain and Chronology, $\chi^2(4) = 20.34$, $p < .001$. There further was an interaction of Chronology and Group, $\chi^2(8) = 23.49$, $p < .01$, and a marginal three-way interaction of Chronology, Domain, and Group, $\chi^2(8) = 14.26$, $p = .075$.

Contrasts showed that there was a significant interaction between Chronology and Domain in group CA, Δ = 19.4 %, $t = 3.48$, $p < .001$, as familiar from Experiment I. This denoted, that *forward* trials in item tasks were responded to more accurately with verbal material than with nonverbal material (Δ = 23.39 %). Within nonverbal material, on the contrary, *backward* trials produced an accuracy benefit (Δ = 13.21 %), see Figure 7.

This interaction effect was absent in first graders (group RA), Δ = 7.89 %, $t = 1.21$, $p = .23$. Instead, in group RA performance on *t1* trials compared to forward or backward trials was significantly lower than in the other groups (all $t > 2.2$, all $p > .05$). This denotes false-alarm errors in trials that might have a specific luring component. Children in first grade more often than all other children seemed to being lured into believing that the second target was from the list too when they had recognized the first one. Figure 7 suggests that this luring effect was especially strong for nonverbal material. However, the interaction with domain did not reach significance.

In group DD, there were no effects of Chronology, $\chi^2(4) = 5.95$, $p = .20$, and no interactions. Also, there were no effects between forward trials in verbal and backward trials in nonverbal material, $\Delta = 4.06\%$, $t = 0.36$, $p = .72$, demonstrating that as a group, children with DD did not benefit from the seriality of this information as the CA group did, and neither were they lured into the false-alarm errors in *t1* trials. Four of the 16 individuals with DD displayed the pattern of forward facilitation in verbal and backward in nonverbal materials. All others showed a main effect of Domain or different patterns.

< Insert Figure 7 about here >

Finally we also analysed whether the use of implicit order information was related to reading, and whether such a relation differed between groups. Adding the separate four covariates for reading to the model on Chronology and Domain, there were different effects across groups.

In group CA, there was the familiar interaction of Chronology and Domain, $\chi^2(4) = 20.5$, $p < .001$, denoting better performance in forward trials for verbal and in backward trials for nonverbal materials. There further were interactions of Chronology with Word Reading Accuracy, $\chi^2(4) = 9.68$, $p < .05$, and with Word Reading Speed, $\chi^2(4) = 11.01$, $p < .05$. These additionally showed that the more correctly and the faster children of grades 2-4 read words, the better they performed on backward trials (accuracy: $t = 2.27$, $p < .05$; syll/sec: $t = 1.99$, $p < .05$) and the less they were lured into false alarm errors in t1 trials (accuracy: $t = 2.04$, $p < .05$; syll/sec: $t = 2.97$, $p < .01$).

As reported above, in group RA there was no interaction between Chronology and Domain, $p > .2$. However, there were interaction effects between Chronology and Word Reading Accuracy, $\chi^2(4) = 14.82$, $p < .01$, and between Chronology and Pseudoword Reading Accuracy, $\chi^2(8) = 18.24$, $p < .01$. Contrasts revealed that the more correctly children in first grade read words and pseudowords, the better they performed in forward (word accuracy: $t = 2.47$, $p < .05$; pseudoword accuracy: $t = 3.32$, $p < .001$) and backward trials (word accuracy: $t = 2.21$, $p < .05$; pseudoword accuracy: $t = 2.47$, $p < .05$).

In group DD, there was no effect of Chronology, $\chi^2(4) < 6$, $p > .2$, and no interactions of either Chronology with Reading (all $\chi^2(4) < 5.6$, all $p > .24$). These differences between groups yielded significant three-way interactions for Chronology, Group, and Word Reading Accuracy, $\chi^2(8) = 24.26$, $p < .01$, and for Chronology, Group, and Pseudoword Reading Accuracy, $\chi^2(8) = 23.5$, $p < .01$. This showed that in the first grade, the use of irrelevant order information during item tasks related to reading accuracy only for the higher spectrum of accurate readers. In the higher grades, all children showed better forward performance for verbal and backward performance for nonverbal materials. Crucially, in children with dyslexia as a group there was no evident recruitment of order information to solve item tasks and no relation of such functions to reading.

Discussion

The second experiment was conducted to investigate serial order and item STM processing in a sample of children with dyslexia compared to the control sample previously reported in Experiment I. Children with dyslexia performed equally well in item tasks as children from the same age group or as younger children matched for reading performance. In serial order tasks, however, children with dyslexia showed lower performance than CA controls and statistically similar performance to the RA group. Although children with dyslexia performed low both in reading and in serial order tasks, they still showed a relation of pseudoword reading speed specifically with verbal order STM. Also, word reading speed related to overall STM performance in group DD. In chronological age matched controls, serial order STM clearly related to both word and pseudoword reading speed, and in the reading age matched control group, to word and pseudoword reading accuracy.

The effects in group DD on first view suggest a delayed development, i.e. describing a group simply of very poor readers. However, it is interesting that the relation between reading and STM measures in group DD was evident for reading speed, similar to the CA group, instead of for reading accuracy, as in the RA group, although reading performance was similar between children with dyslexia and the RA group. In this contrast, there might have been an influence of a *t1* luring effect leading to weaker item task performance in group RA. These children might have tried to shorten the item tasks by answering *yes* when only the first probe had been in the list, committing an error. However, doing so, they still used the serial order of the probes, while there was no such effect in group DD. Also, children of group RA showed less of a *t1* luring effect the better they could read words. These considerations need to be taken into account although the influence of item task performance for the contrast in question is small. The focus on reading accuracy versus speed in first grade is well known in the literature (Aro & Wimmer, 2003; Landerl & Wimmer, 2008; Wimmer & Mayringer, 2002) and we suggest that our results speak more for a rejection of the gap hypothesis than its adoption. The scaling of the reading test, which was shorter for children in grade 1 than all other participants, might also have influenced the interaction of reading and serial order STM in group RA, see the Study limitations section. This taps the magnitude, but not the quality of the interaction effect of reading and serial order STM in group RA.

One tentative suggestion might be that children with dyslexia have already progressed with their reading strategies similar to chronological age matched peers, however, without fully acquiring the necessary phonological and grapheme knowledge for the word recognition skills to proceed. Furthermore, the control group children demonstrated a facilitation effect during item tasks by using the inherent seriality of information within the item tasks to their benefit (see Figure 7). This was presumably done implicitly since participants were not instructed to do so. As would be expected in shallow orthographies, this order facilitation effect was also related to reading accuracy in group RA and to reading speed measures in group CA. Importantly, in children with dyslexia there were no effects of recruiting chronological target order during item tasks as well as no such relations to reading performance, once again demonstrating that children with DD display a deficit or unavailability in their ability to process serial order information compared to controls. Please note that these analyses on target order in item tasks are run on small cell sizes. We suggest that they corroborate our evidence rather than representing strong conclusive results on their own.

It should also be noted that pseudoword reading accuracy showed no relation to STM performance in Experiment I, but it did so in Experiment II. This may be due to the different number of factor levels in Grade and Group. Note that we report only results that reach significance in models that control for all the factors and effects relevant for the hypotheses. The variance distribution in the analysis by Grade may have led to lower significance in Experiment I. In the slightly coarser analysis by Group in Experiment II, also pseudoword reading related to serial order task performance.

In comprehension, our data strongly suggest a relationship of serial order STM abilities and reading skills that is most pronounced during the time of application and training of GPC decoding skills in a way in which both skills support each other. This supposed co-development seems to be hampered in children with dyslexia to an extent that does not allow automatization of reading as in other children. In addition, children with DD seemed not to benefit from the implicit seriality embedded within the item tasks as the CA group did. For these children, serial order processing may therefore be confined to more conscious and effortful processing than in chronological age matched controls who recruit serial order information spontaneously and implicitly. Even high-performing first graders showed a similar spontaneous use of order information in item tasks.

In a study with dyslexic adults and normal readers, Romani, Tsouknida, and Olson (2015b) differentiated between temporal order in single presentation, spatial order in parallel presentation, and consolidation of serial information, which represent different components of the spectrum of serial order tasks. In accordance with our results, they showed that different serial order components are related to distinct skills of orthographic processing, and also, that adults with and without dyslexia showed different patterns of relations between serial order components, reading, and orthography. Taken together, this emphasizes the need for future studies to investigate compensatory or strategically different processing methods for children who struggle with reading and who show deficits in serial order processing. Overall, our results and others (Bogaerts et al., 2016; Bogaerts, Szmalec, Hachmann, Page, & Duyck, 2015; Cowan et al., 2017; Hachmann et al., 2014; Majerus et al., 2009; Martinez Perez, Majerus, Mahot, et al., 2012; Martinez Perez, Majerus, & Poncelet, 2012; Martinez Perez et al., 2013; Romani et al., 2015b) demonstrate how deficits in serial order abilities can repeatedly be found in poor readers; but see also (Staels & Van den Broeck, 2013), who could not find similar impairments. The majority of results suggest a co-developing relationship between these processes that is hampered in poor readers and individuals with dyslexia.

General discussion

Reading development relies on building connections between abstract visual graphical images and phonological word forms embedded within the serial order of consecutive information. It has been proposed that cognition (Shing et al., 2010) and reading in particular (Klimesch, 2012) are fundamentally based on the intrinsic processing and development of serial order functions (for a summary see Klimesch, Sauseng, & Hanslmayr, 2007). Recently, this proposal has received renewed attention and been more substantially investigated to elucidate the relation of serial order processing to language development (Majerus & Boukebza, 2013; Majerus, Poncelet, Greffe, et al., 2006; Majerus et al., 2008), to reading (Binamé & Poncelet, 2016; Bogaerts et al., 2016; Martinez Perez, Majerus, & Poncelet, 2012), and to developmental dyslexia (Bogaerts, Szmalec, Hachmann, Page, & Duyck, 2015; Cowan et al., 2017; Hachmann et al., 2014; Howard et al., 2006; Martinez Perez, Majerus, Mahot, et al., 2012; Martinez Perez et al., 2013; Romani et al., 2015b). This was partially due to new experimental paradigms and comprehensive models of serial order processing within short-term memory (Brown et al., 2000; Gupta et al., 2005; Majerus, Poncelet, Greffe, et al., 2006; Page & Norris, 2009). In our current study, we created a novel experimental design that balanced order and item STM tasks as well as manipulated the requirements of verbal and nonverbal material as originating from the same task using the example from Hsieh et al. (2011). Experiment I was conducted to investigate STM for serial order versus identity information in relation to reading development in elementary school children. In a second experiment, we administered the same tasks to children with developmental dyslexia (DD) compared to the chronological age matched control group (CA) and reading age matched control group (RA) from Experiment I. It is important to note that this particular series of experiments were a cross-sectional design and therefore any definitive conclusions related to continuous cognitive development of reading skills and STM abilities in children will need to be reserved until follow-up longitudinal studies can be conducted using similar techniques. However with this caveat in mind, our current results strongly suggest a developmental relationship between children's reading abilities and serial order STM performance, and generalized STM deficits in maintaining serial order information in DD.

Results from the first experiment suggest a close relationship of STM for serial order and reading skills that showed a boost in grade 3. This relationship was further most pronounced for reading accuracy in first grade children, while it was evident for reading speed measures in grades 2 to 4. In addition, also the spontaneous recruitment of serial order information during item task execution was related to reading abilities during these grades in a similar way.

In Experiment II, we found that children with DD displayed marked deficits for serial order tasks for both verbal and non-verbal material compared to item tasks (see Figure 6). This was also true when compared to their chronological age matched peers.

On first view, the differences between children with dyslexia and both control groups suggest that children in group DD performed merely like very poor readers, supporting the gap hypothesis of dyslexia. We found substantial differences between groups that went beyond a gap hypothesis; however, they are not conclusive enough to fully reject the gap hypothesis. There was a difference in the quality of the interaction between reading and serial order STM in group DD as compared to group RA: the relations that were still evident in the albeit low performance of serial order STM and reading in group DD tapped reading speed as in group CA, not reading accuracy as in first grade (RA). Also, children with dyslexia showed no sign of recruiting implicit serial order information during item tasks as did RA controls (see Figure 7). We therefore remain in favour of a rejection of the gap hypothesis according to our results, but further investigation is indicated also concerning the type of reading test and possible sampling, see the Study limitation section for details.

In children with dyslexia, both reading and serial order STM skills seem to remain behind that of their age matched peers, although over time they start to adopt compensatory age-adequate reading strategies. This may be occurring despite a lack of accurate grapheme and word recognition skills. We therefore conclude that initial problems to process serial order in STM can lead to an exacerbated lack of the co-development of serial order memory processes with reading skills that could finally lead to differences in developmental trajectories. However, similar to the discussions of the role of a phonological deficit in dyslexia (Morais & Kolinsky, 1994; Pennington, 2006; Ramus & Szenkovits, 2008), we sustain that also serial order STM can be impaired as a result of a lack of reading experience and training. In our view, serial order STM - just as a phonological deficit - don't necessarily causes a reading deficit, but it may go along with it and therefore represents a useful measure for the characterisation and treatment of dyslexia.

Study limitations

Children of the control groups went to two different schools. The school and its teaching philosophy arguably could have had an influence on memory and reading performance. To control for this possibility, we ran a separate analysis for Experiment I of the model on Grade and the four models in which each reading measure was added separately, by including a random intercept term for Grade as nested in School. There were no differences in the significance of effects through all five models. Unfortunately, and this has to be considered a limiting factor of our results, we could not retrieve more fine-grained information on teaching clusters like the classroom of each individual. There were about sixteen classrooms involved (about eight per school by two per grade) and each had their own pace and method variety within a rather strict school specific teaching plan.

Another limiting factor concerns the scaling of the reading tests. This is a universal problem when testing at the very low end of performance. In this study, all children from grades 2-4 performed the entire reading test for word and pseudowords reading. This was too demanding for first graders, so we asked them to read only the first page of each test. To maintain scaling, we used raw error rates and syllables per second for all children in all analyses. Since the test had to be coordinated also with the therapy centre's testing metric, to our knowledge this was the best possible solution. But we are conscious of the problem that scores in grade 1 might have been less influenced by stamina and proactive interferences than in higher grades. As support for the credibility of our data we might put forward that artificially boosted reading performance in first grade would work against our hypothesis of an increase in the interaction between serial order STM and reading performance from grades 2-3. As there were already substantial serial order STM skills in grade 1 there was a better chance for this interaction to be visible when reading performance was higher. Children of group DD, on the contrary, mostly performed the whole reading test, which is especially exhausting for them. So the scaling of the reading measures might have influenced the comparison between this interaction in group RA and in group DD in Experiment II, questioning the possible rejection of a gap hypothesis. However, since serial order STM skills were lower in group DD and their interaction with reading concerned reading speed, not accuracy as in group RA, we favour a rejection of the gap hypothesis according to our data. But with the use of novel experimental paradigms such as here and due to the heterogeneity of children's cognitive development across samples and the difficult nature of matching the relative trajectories of these developments between groups, we strongly emphasize the importance of further systematic investigations hereof.

Verbal item knowledge and serial order STM

Phonological awareness tasks often include encoding and manipulation of linguistic units in serial order, which might trigger an added difficulty for children with poor reading ability: when assuming that the co-development between order STM and reading skills is hampered, these children could possibly be faced with more effortful serial order processing requirements plus the resulting untrained linguistic item knowledge (phonemes, graphemes, and words).

In accordance with our results presented here, impaired serial order performance in children with dyslexia has also been found in a study by Martinez Perez, Majerus, Mahot, et al. (2012), where impaired serial order STM performance was found in DD when compared to reading age and chronological age matched controls. In their study, however, Martinez Perez and colleagues found differences between children with dyslexia and chronological age matched controls also in item task performance. There are two important differences between our studies which suggest that verbal item impairments in dyslexia depend on the task and material that is used. In a previous study, Hachmann and colleagues (2014) found that dyslexic adults performed at the same level with controls in verbal and nonverbal item tasks using a recognition task with images and words, similar to the one used here, but the task was not specifically balanced to control for performance differences as was done in our current study. Martinez Perez, Majerus, Mahot, et al. (2012) and Martinez Perez et al. (2013) used repetition of very short pseudowords as the verbal item task and found impaired performance of dyslexic individuals. First, it may be argued that pseudoword processing requires residual serial order maintenance in STM (for a discussion on task selection see also Hsieh et al., 2011; Romani, Tsouknida, & Olson, 2015a). Pseudowords need to be decoded sequentially, even more so since accommodating semantic and lexical support is missing (Ziegler & Goswami, 2005). Second, as Majerus and Cowan (2016) argue correctly, our item task measure most likely tap into STM ability on the level of whole words. Children with reading problems, though, are reported to struggle most on the sublexical level, particularly in grapheme-phoneme conversions (GPC) (Ramus & Szenkovits, 2008). However, GPC item skills are thought to develop along with reading exposure, and hence can be impaired as a consequence of reading problems rather than a cause (Castles & Coltheart, 2004; Morais &

Kolinsky, 1994; Pattamadilok, Lafontaine, Morais, & Kolinsky, 2010; Ventura, Kolinsky, Querido, Fernandes, & Morais, 2007). We therefore purposefully chose the word level as our measure to address as little co-affected item knowledge as possible when assessing serial order memory in dyslexia. Other than verbal item knowledge, that should be relatively easy to acquire in a transparent language such as Italian (Barca, Burani, Di Filippo, & Zoccolotti, 2006), serial order memory is conceptualized as a domain-general and separate function (Burgess & Hitch, 1999; Majerus et al., 2008; Page & Norris, 2009), and should therefore be an independent albeit influential factor in reading acquisition across languages. Impairments in serial order processing are reflected in alterations to the neural network for domain-general serial order STM processing in adults with developmental dyslexia, as shown by Basar (2013). Overall, this emphasizes the importance of using item tasks that minimize the requirement of secondarily impaired long-term knowledge when looking into the hypothesized differences between order and item functions in STM and their relations to reading in dyslexia and controls. Beyond this contrast it is clearly important to deeply investigate the exact item impairments of individuals with dyslexia for therapeutic, practical and theoretical reasons.

Conclusion

We investigated the relationship of short-term memory (STM) for ordered information and reading abilities in elementary school children from grades 1 through 4 using a novel balanced double-probe paradigm including verbal and nonverbal material. Results show that STM abilities for ordered information were significantly related to reading abilities throughout all grade levels, even for nonverbal STM tasks. The strongest relationship was evident between grades 2 and 3, during which learning is thought to be the most dependent upon grapheme-phoneme conversion (Share, 2004). In Experiment II, we also assessed serial order short-term memory performance in children with developmental dyslexia (DD) and found they displayed specific deficits in serial order STM compared to chronological age matched controls. Furthermore, the control group implicitly recruited sequential information to facilitate item memory performance, but this facilitation was absent in children with DD.

Overall, these results demonstrate that a domain-general STM function for serial order information is strongly related with children's reading abilities and this particular relationship is markedly impaired in children with DD. Furthermore, our results suggest that other than beginning readers of the reading-age matched control group, who showed initial relationships between reading and serial order STM, children with DD struggle with this co-development and eventually proceed in their reading strategies with limited underlying STM abilities to maintain serially ordered information, which puts them at considerable disadvantage. The implications of this interpretation should be tested further with longitudinal studies.

The serial order approach may also be fruitful for diagnosis and remediation of dyslexia. The STM for serial order information may be trainable to obtain a ripple effect on children's reading skills. Also, serial order STM impairments provide a valid additional explanation for poor performance in phonological and other meta-linguistic knowledge tests that introduce serial order processing requirements. Applying the distinction between difficulties with specific verbal item knowledge and general serial order requirements, diagnosis and therapy can be targeted much more individually to support children who struggle with reading.

References

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological bulletin, 135*(1), 50-68.

Ahissar, M., Lubin, Y., Putter-Katz, H., & Banai, K. (2006). Dyslexia and the failure to form a perceptual anchor. *Nature neuroscience, 9*(12), 1558-1564.

Aro, M., & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics, 24*(04).

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews neuroscience, 4*(10), 829-839.

Banai, K., & Ahissar, M. (2010). On the importance of anchoring and the consequences of its impairment in dyslexia. *Dyslexia, 16*(3), 240-257.

Barca, L., Burani, C., Di Filippo, G., & Zoccolotti, P. (2006). Italian developmental dyslexic and proficient readers: where are the differences? *Brain and Language, 98*(3), 347-351.

Basar, E. (2013). Brain oscillations in neuropsychiatric desease. *Dialogues in Clinical Neuroscience, 15*(3), 291-300.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R., & Singmann, H. (2015). lme4: Linear mixed-effects models using Eigen and S4 *R package version 1.1-10*: GPL.

Binamé, F., & Poncelet, M. (2016). The development of the abilities to acquire novel detailed orthographic representations and maintain them in long-term memory. *J Exp Child Psychol, 143*, 14-33.

Bogaerts, L., Szmalec, A., De Maeyer, M., Page, M. P., & Duyck, W. (2016). The involvement of long-term serial-order memory in reading development: A longitudinal study. *J Exp Child Psychol, 145*, 139-156.

Bogaerts, L., Szmalec, A., Hachmann, W. M., Page, M. P., & Duyck, W. (2015). Linking memory and language: Evidence for a serial-order learning impairment in dyslexia. *Research in Developmental Disabilities, 43*, 106-122.

Bogaerts, L., Szmalec, A., Hachmann, W. M., Page, M. P., Woumans, E., & Duyck, W. (2015). Increased susceptibility to proactive interference in adults with dyslexia? *Memory, 23*(2), 268-277.

Brown, G. D. A., Preese, T., & Hulme, C. (2000). Oscillator-Based Memory for Serial Order. *Psychological review, 107*(1), 127-181.

Burgess, N., & Hitch, G. J. (1999). Memory for Serial Order: A Network Model of the Phonological Loop and Its Timing. *Psychological review, 106*(3), 551-581.

Byrne, B. (1981). Reading disability, linguistic access and short-term memory: Comments prompted by Jorm's review of developmental dyslexia. *Australian Journal of Psychology, 33*(1), 83-95. doi: 10.1080/00049538108255294

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition, 91*(1), 77-111.

Cohen, R. L., Netley, C., & Clarke, M. A. (1984). On the Generality of the Short-Term Memory/Reading Ability Relationship. *Journal of learning disabilities, 17*(4), 218-221.

Cowan, N., Hogan, T. P., Alt, M., Green, S., Cabbage, K. L., Brinkley, S., & Gray, S. (2017). Short-term Memory in Childhood Dyslexia: Deficient Serial Order in Multiple Modalities. *Dyslexia, 23*(3), 209-233. doi: 10.1002/dys.1557

Engle, R. W., & Mobley, L. A. (1976). The modality effect: What happens in long-term memory? *Journal of Verbal Learning and Verbal Behavior, 15*(5), 519-527.

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., . . . Graves, S. (2016). Package 'car': Companion to Applied Regression: GPL<=2.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for Visual Word Recognition and Orthographical Depth: A Multilingual Comparison. *Journal of Experimental Psychology: Human perception and performance, 13*(1), 104-115.

Gervain, J., Nespor, M., Mazuka, R., Horie, R., & Mehler, J. (2008). Bootstrapping word order in prelexical infants: a Japanese-Italian cross-linguistic study. *Cognitive Psychology, 57*(1), 56-74.

Gouldthorp, B., Katsipis, L., & Mueller, C. (2017). An Investigation of the Role of Sequencing in Children's Reading Comprehension. *Reading Research Quarterly*. doi: 10.1002/rrq.186

Gupta, P., Lipinski, J., & Aktunc, E. (2005). Reexamining the phonological similarity effect in immediate serial recall: The roles of type of similarity, category cuing, and item recall. *Memory and Cognition, 33*(6), 1001-1016.

Hachmann, W. M., Bogaerts, L., Szmalec, A., Woumans, E., Duyck, W., & Job, R. (2014). Short-term memory for order but not for item information is impaired in developmental dyslexia. *Ann Dyslexia, 64*(2), 121-136. doi: 10.1007/s11881-013-0089-5

Hothorn, T., Bretz, F., Westfall, P., Heiberger, R. M., Schuetzenmeister, A., Scheibe, S., & Hothorn, M. T. (2016). Package 'multcomp': Simultaneous Inference in General Parametric Models: GPL.

Houghton, G., & Hartley, T. (1995). Parallel models of serial behavior - Lashley revisited. *Psyche, 2*(25), 1-25.

Howard, J. H., Jr., Howard, D. V., Japikse, K. C., & Eden, G. F. (2006). Dyslexics are impaired on implicit higher-order sequence learning, but not on implicit spatial context learning. *Neuropsychologia, 44*(7), 1131-1144.

Hsieh, L.-T., Ekstrom, A. D., & Ranganath, C. (2011). Neural oscillations associated with item and temporal order maintenance in working memory. *The Journal of Neuroscience, 31*(30), 10803-10810.

Janssen, D. P. (2012). Twice random, once mixed: Applying mixed models to simultaneously analyze random effects of language and participants. *Behavior Research Methods, 44*(1), 232-247. doi: 10.3758/s13428-011-0145-1

Katz, L., & Frost, R. (1992). The Reading Process is Different for Different Orthographies: The Orthographic Depth Hypothesis. In R. Frost & L. Katz (Eds.), *Advances in Psychology: Orthography, Phonology, Morphology, and Meaning* (Vol. 94, pp. 67-84). Amsterdam: Elsevier Science Publishers B.V.

Klimesch, W. (2012). alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Science, 16*(12), 606-617.

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Research Review, 53*(1), 63-88.

Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology, 100*(1), 150-161. doi: 10.1037/0022-0663.100.1.150

Leclercq, A.-L., & Majerus, S. (2010). Serial-order short-term memory predicts vocabulary development: evidence from a longitudinal study. *Dev Psychol, 46*(2), 417-427. doi: 10.1037/a0018540

Leong, V., & Goswami, U. (2014). Impaired extraction of speech rhythm from temporal modulation patterns in speech in developmental dyslexia. *Frontiers in human neuroscience, 8*, 96. doi: 10.3389/fnhum.2014.00096

Majerus, S., & Boukebza, C. (2013). Short-term memory for serial order supports vocabulary development: new evidence from a novel word learning paradigm. *J Exp Child Psychol, 116*(4), 811-828. doi: 10.1016/j.jecp.2013.07.014

Majerus, S., & Cowan, N. (2016). The Nature of Verbal Short-Term Impairment in Dyslexia: The Importance of Serial Order. *Frontiers in Psychology, 7*, 1522. doi: 10.3389/fpsyg.2016.01522

Majerus, S., D□Argembeau, A., Martinez Perez, T., Belayachi, S., Van der Linden, M., Collette, F., . . . Maquet, P. (2010). The Commonality of Neural Networks for Verbal and Visual Short-term Memory. *Journal of cognitive neuroscience, 22*(11), 2570-2593. doi: 10.1162/jocn.2009.21378

Majerus, S., Leclercq, A.-L., Grossmann, A., Billard, C., Touzin, M., Van Der Linden, M., & Poncelet, M. (2009). Serial order short-term memory capacities and specific language impairment: no evidence for a causal association. *Cortex, 45*(6), 708-720. doi: 10.1016/j.cortex.2008.10.006

Majerus, S., Poncelet, M., Elsen, B., & van der Linden, M. (2006). Exploring the relationship between new word learning and short-term memory for serial order recall, item recall, and item recognition. *European Journal of Cognitive Psychology, 18*(6), 848-873. doi: 10.1080/09541440500446476

Majerus, S., Poncelet, M., Greffe, C., & Van Der Linden, M. (2006). Relations between vocabulary development and verbal short-term memory: The relative importance of short-term memory for serial order and item information. *J Exp Child Psychol, 93*(2), 95-119. doi: 10.1016/j.jecp.2005.07.005

Majerus, S., Poncelet, M., Van Der Linden, M., & Weekes, B. S. (2008). Lexical learning in bilingual adults: the relative importance of short-term memory for serial order and phonological knowledge. *Cognition, 107*(2), 395-419. doi: 10.1016/j.cognition.2007.10.003

Majerus, S., Van der Linden, M., Mulder, L., Meulemans, T., & Peters, F. (2004). Verbal short-term memory reflects the sublexical organization of the phonological language network: Evidence from an incidental phonotactic learning paradigm. *Journal of Memory and Language, 51*(2), 297-306. doi: 10.1016/j.jml.2004.05.002

Martín, J. A. M., & Pérez, M. E. G. (2008). ONESC: A database of orthographic neighbors for Spanish read by children. *Behavior Research Methods, 40*(1), 191-197.

Martinez Perez, T., Majerus, S., Mahot, A., & Poncelet, M. (2012). Evidence for a specific impairment of serial order short-term memory in dyslexic children. *Dyslexia, 18*(2), 94-109. doi: 10.1002/dys.1438

Martinez Perez, T., Majerus, S., & Poncelet, M. (2012). The contribution of short-term memory for serial order to early reading acquisition: Evidence from a longitudinal study. *J Exp Child Psychol, 111*(4), 708-723. doi: 10.1016/j.jecp.2011.11.007

Martinez Perez, T., Majerus, S., & Poncelet, M. (2013). Impaired short-term memory for order in adults with dyslexia. *Research in Developmental Disabilities, 34*(7), 2211-2223.

Morais, J., & Kolinsky, R. (1994). Perception and awareness in phonological processing: the case of the phoneme. *Cognition, 50*, 287-297. doi: 10.1016/0010-0277(94)90032-9

Nisi, M., Longoni, A. M., & Snodgrass, J. G. (2000). Misure italiane per l'accordo sul nome, familiarità ed età di acquisizione per le 260 figure di snodgrass e vanderwart (1980). *Giornale Italiano Di Psicologia, 27*(1), 205-218. doi: 10.1421/288

Nithart, C., Demont, E., Majerus, S., Leybaert, J., Poncelet, M., & Metz-Lutz, M.-N. (2009). Reading disabilities in SLI and dyslexia result from distinct phonological impairments. *Developmental neuropsychology, 34*(3), 296-311. doi: 10.1080/87565640902801841

Page, M. P., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 364*(1536), 3737-3753. doi: 10.1098/rstb.2009.0173

Pagliarini, E., Guasti, M. T., Toneatto, C., Granocchio, E., Riva, F., Sarti, D., . . . Stucchi, N. (2015). Dyslexic children fail to comply with the rhythmic constraints of handwriting. *Human Movement Science, 42*, 161-182. doi: 10.1016/j.humov.2015.04.012

Pattamadilok, C., Lafontaine, H., Morais, J., & Kolinsky, R. (2010). Auditory Word Serial Recall Benefits from Orthographic Dissimilarity. *Language and Speech, 53*(3), 321-341. doi: 10.1177/0023830910371450

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition, 101*(2), 385-413. doi: 10.1016/j.cognition.2006.04.008

Poncelet, M., Schyns, T., & Majerus, S. (2003). Further evidence for persisting difficulties in orthographic learning in highly educated adults with a history of developmental dyslexia. *Brain and Language, 87*(1), 145-146. doi: 10.1016/S0093-934X(03)00241-4

R Core Team, R. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2015. URL h ttp. GPL: www. R-project. org.

Ramus, F., & Szenkovits, G. (2008). What phonological deficit? *The Quarterly Journal of Experimental Psychology, 61*(1), 129-141. doi: 10.1080/17470210701508822

Romani, C., Tsouknida, E., & Olson, A. (2015a). Encoding order and developmental dyslexia: a family of skills predicting different orthographic components. *The Quarterly Journal of Experimental Psychology, 68*(1), 99-128. doi: 10.1080/17470218.2014.938666

Romani, C., Tsouknida, E., & Olson, A. (2015b). Encoding order and developmental dyslexia: a family of skills predicting different orthographic components. *Q J Exp Psychol (Hove), 68*(1), 99-128.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception, 33*(2), 217-236. doi: 10.1068/p5117

Sartori, G., Job, R., & Tressoldi, P. E. (2007). *DDE-2: Batteria per la valutazione della dislessia e della disortografia evolutiva (Battery for the assessment of developmental dyslexia and dysorthographia)*. Firenze: Giunti OS.

Schuck, N., Gaschler, R., & Frensch, P. (2012). Implicit learning of what comes when and where within a sequence: The time-course of acquiring serial position-item and item-item associations to represent serial order. *Advances in Cognitive Psychology, 8*(2), 83-97. doi: 10.5709/acp-0106-0

Share, D. L. (2004). Orthographic learning at a glance: on the time course and developmental onset of self-teaching. *J Exp Child Psychol, 87*(4), 267-298. doi: 10.1016/j.jecp.2004.01.001

Shing, Y. L., Werkle-Bergner, M., Brehmer, Y., Muller, V., Li, S.-C., & Lindenberger, U. (2010). Episodic memory across the lifespan: the contributions of associative and strategic components. *Neuroscience and Biobehavioral Review, 34*(7), 1080-1091.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception, 28*(9), 1059-1074.

Smith-Spark, J. H., & Fisk, J. E. (2007). Working memory functioning in developmental dyslexia. *Memory, 15*(1), 34-56. doi: 10.1080/09658210601043384

Snowling, M. J., & Hulme, C. (2005). *The Science of Reading: A Handbook*. Oxford: Blackwell Publishing Ltd.

Solomon, P., Kubzansky, P. E., Leiderman, P. H., Mendelson, J. H., Trumbull, R., & Wexler, D. (1961). Sensory deprivation.

Staels, E., & Van den Broeck, W. (2013). Order short-term memory is not impaired in dyslexia and does not affect orthographic learning. *Frontiers in human neuroscience, 8*, 732-732. doi: 10.3389/fnhum.2014.00732

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature reviews neuroscience, 9*(4), 255-266. doi: 10.1038/nrn2331

Szmalec, A., Loncke, M., Page, M. P., & Duyck, W. (2011). Order or Disorder? Impaired Hebb Learning in Dyslexia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(5), 1270–1279.

Tomblin, J. B., Mainela-Arnold, E., & Zhang, X. (2007). Procedural learning in adolescents with and without specific language impairment. *Language Learning and Development, 3*(4), 269-293.

Ventura, P., Kolinsky, R., Querido, J.-L., Fernandes, S., & Morais, J. (2007). Is phonological encoding in naming influenced by literacy? *Journal of Psycholinguistic Research, 36*(5), 341-360.

Wechsler, D. (1991). *WISC-III: Wechsler intelligence scale for children*: Psychological Corporation.

Wimmer, H., & Mayringer, H. (2002). Dysfluent reading in the absence of spelling difficulties: A specific disability in regular orthographies. *Journal of Educational Psychology, 94*(2), 272-277.

Ziegler, J. C., Bertrand, D., Leté, B., & Grainger, J. (2014). Orthographic and phonological contributions to reading development: tracking developmental trajectories using masked priming. *Dev Psychol, 50*(4), 1026-1036. doi: 10.1037/a0035187

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological bulletin, 131*(1), 3-29.

Zoccolotti, P., De Luca, M., Di Filippo, G., Judica, A., & Martelli, M. (2009). Reading development in an orthographically regular language: effects of length, frequency, lexicality and global processing ability. *Reading and Writing, 22*(9), 1053-1079.

Figure 1: Design of the four experimental tasks: Serial order and item conditions are crossed with verbal and nonverbal material. In item tasks, participants judge whether both target items had indeed been part of the previous list or not, and in order tasks, participants decide whether the relative order of the two targets is the same as during list presentation.

Figure 2: Percentage of correct responses for each of the four STM tasks by Grade. Polygons denote standard errors.

Figure 3a: STM performance by Word Reading Accuracy for each grade. Betas represent estimates of single contrasts within the respective mixed effects regression.

Figure 3b: STM performance by Word Reading Speed for each grade.

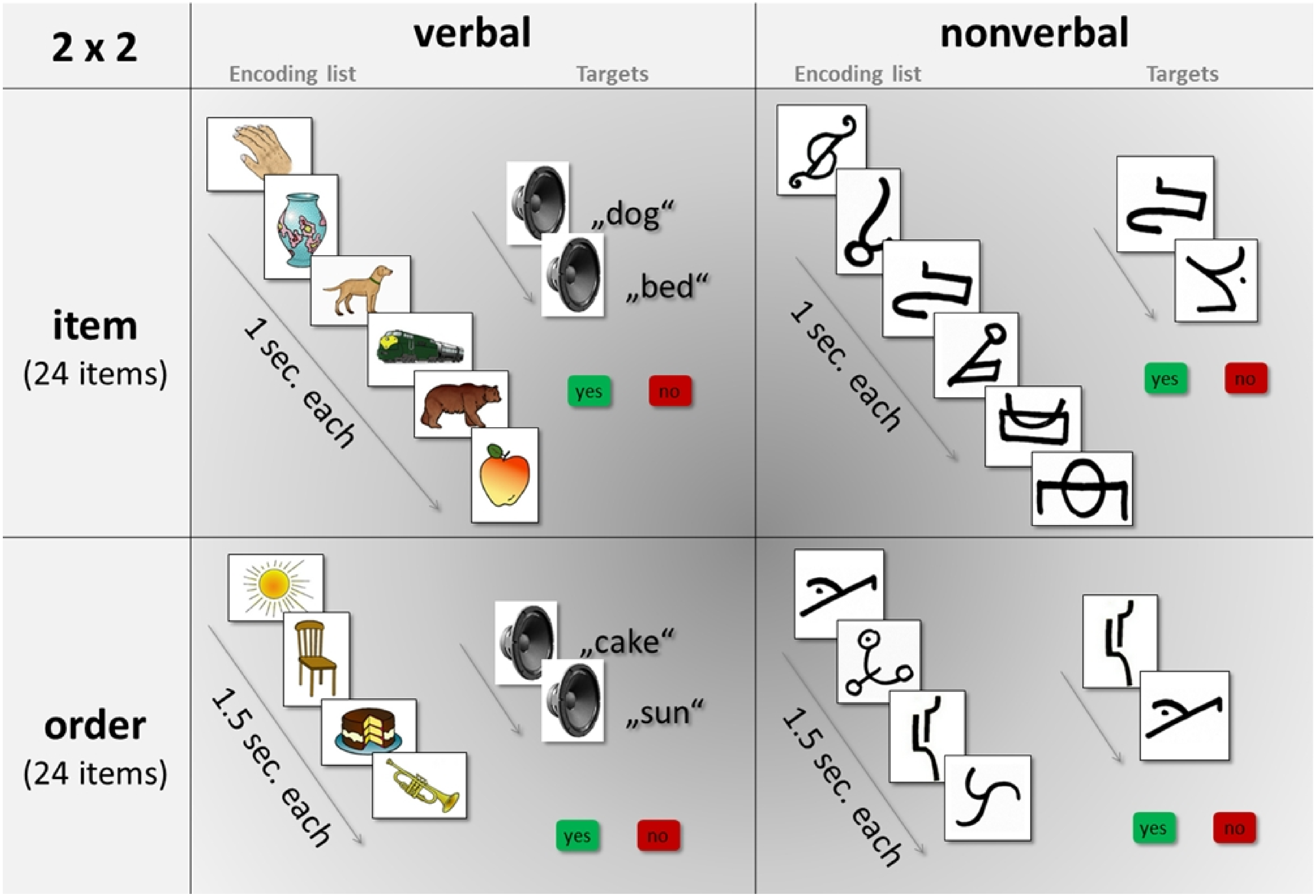
Figure 4a: STM performance by Pseudoword Reading Accuracy for each grade. Betas represent estimates of single contrasts within the respective mixed effects regression.

Figure 4b: STM performance by Pseudoword Reading Speed for each grade.

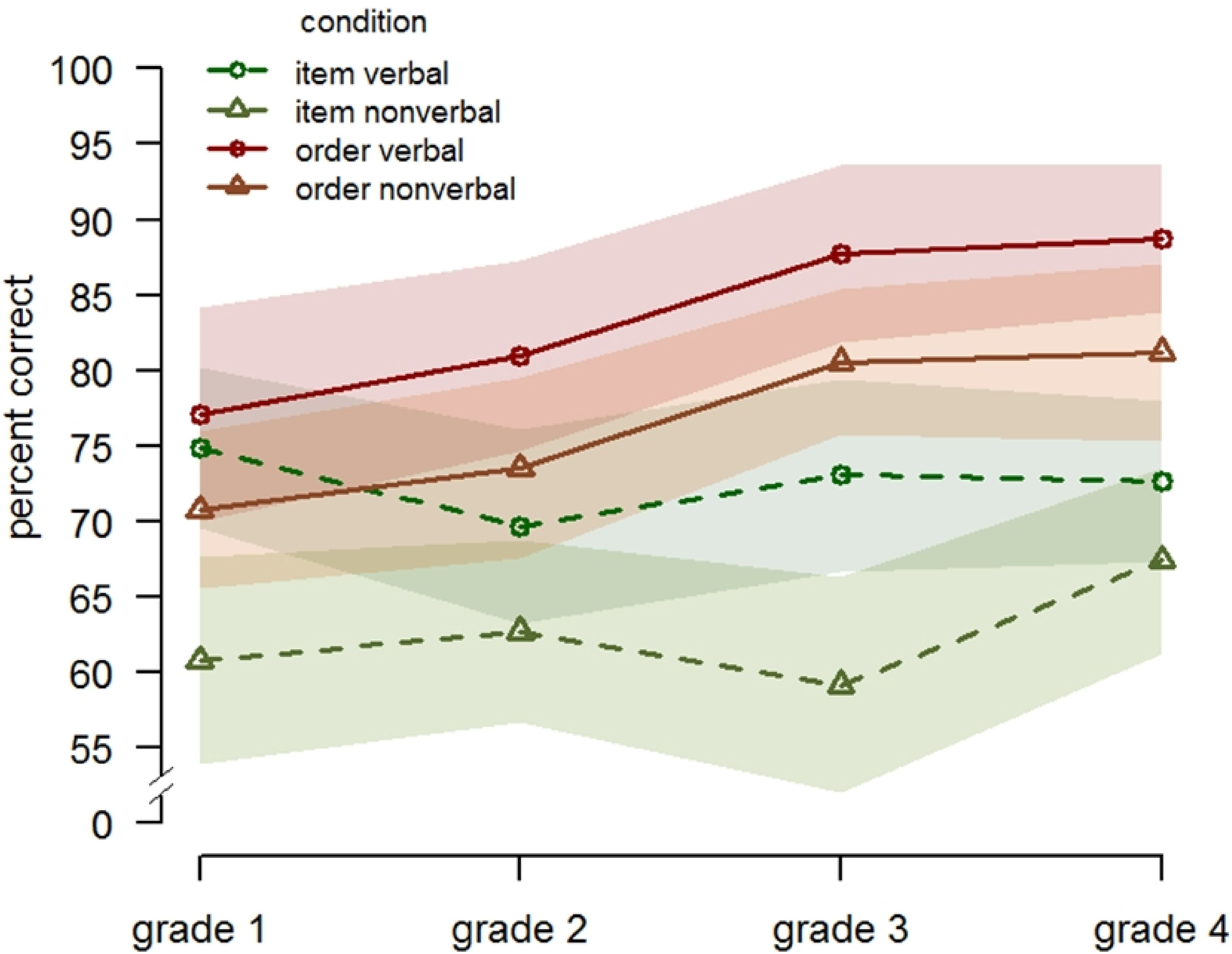
Figure 5: Percentage of correct responses in item tasks by Chronology of target presentation for verbal and nonverbal material. Whiskers denote standard errors. * *fw*: both probes were in the list in same order; *bw*: both probes were in the list in opposite order; *t1*: only first probe was in the list; *t2*: only second probe was in the list; *none*: neither probe was in the list.

Figure 6: Bar plot of STM performance by Group, Domain, and Task. Whiskers denote standard errors.

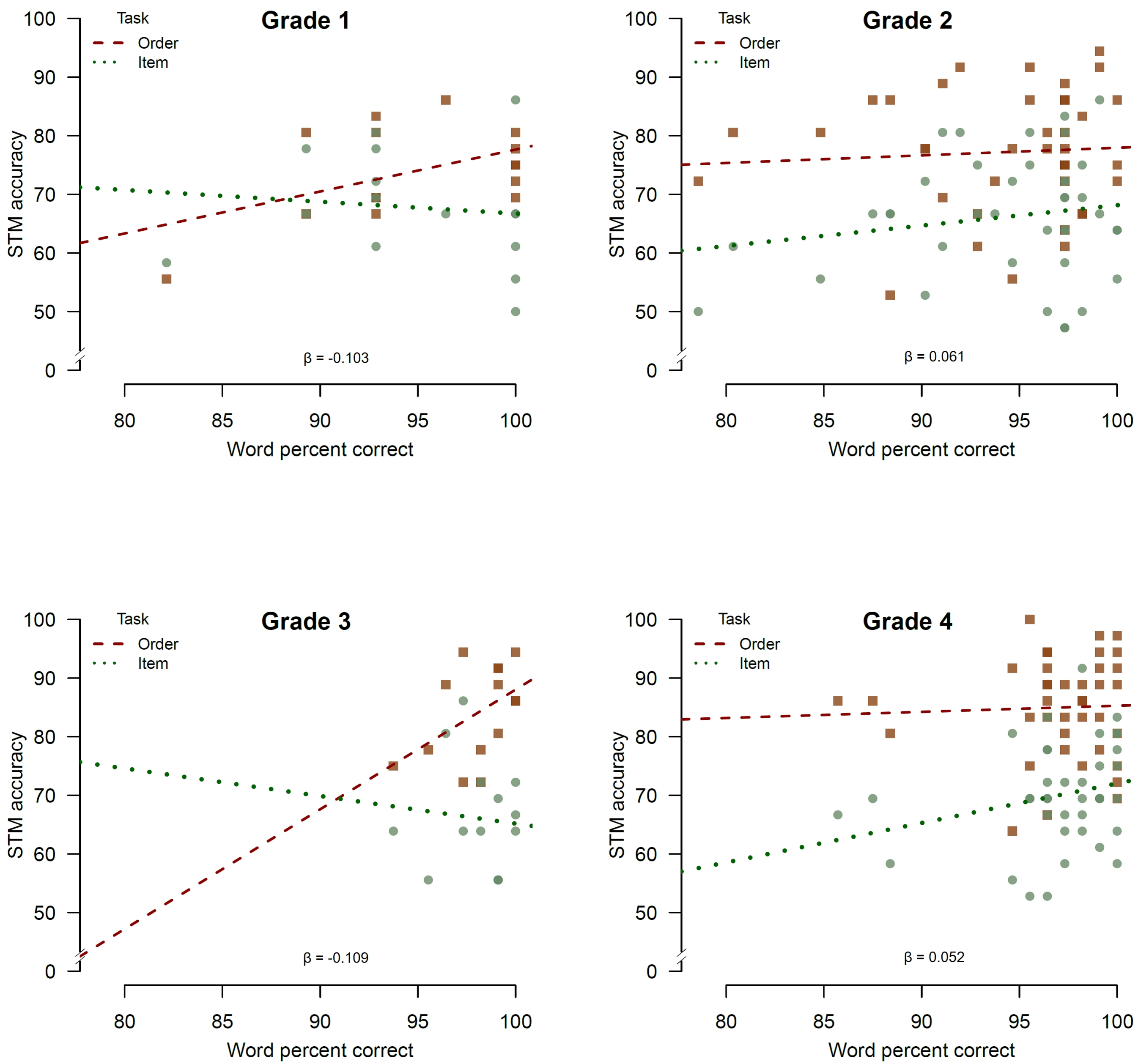
Figure 7: Percentage of correct responses in item tasks within each group by Chronology of target presentation with verbal and nonverbal material. Whiskers denote standard errors. **forward*: both probes were in the list in same order; *backward*: both probes were in the list in opposite order; *target 1*: only first probe was in the list; *target 2*: only second probe was in the list; *none*: neither probe was in the list.



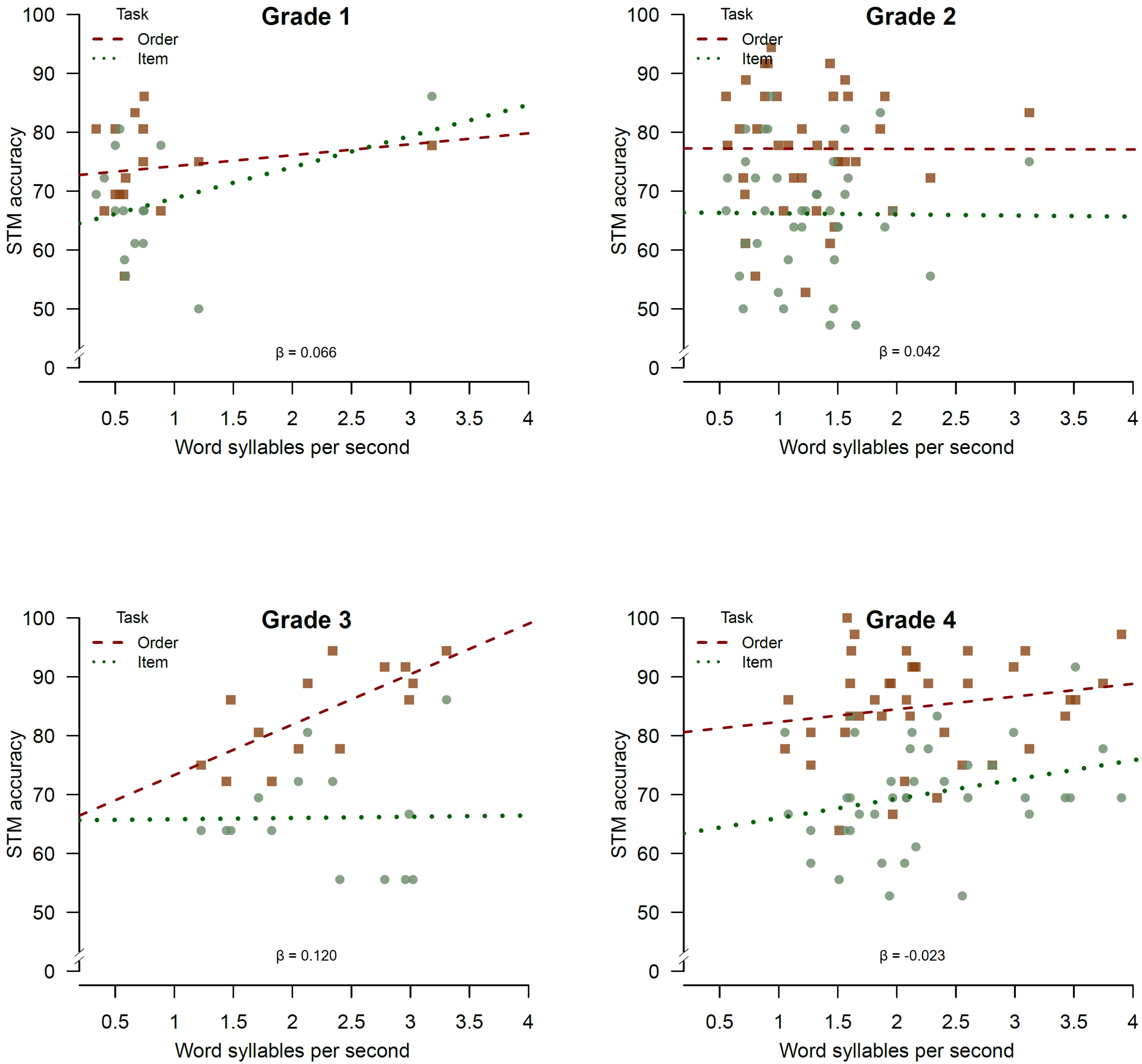
2 x 2
verbal
Encoding list
Targets
nonverbal
Encoding list
Targets
item
(24 items)
1 sec. each
„dog“
„bed“
yes
no
1 sec. each
yes
no
order
(24 items)
1.5 sec. each
„cake“
„sun“
yes
no
1.5 sec. each
yes
no

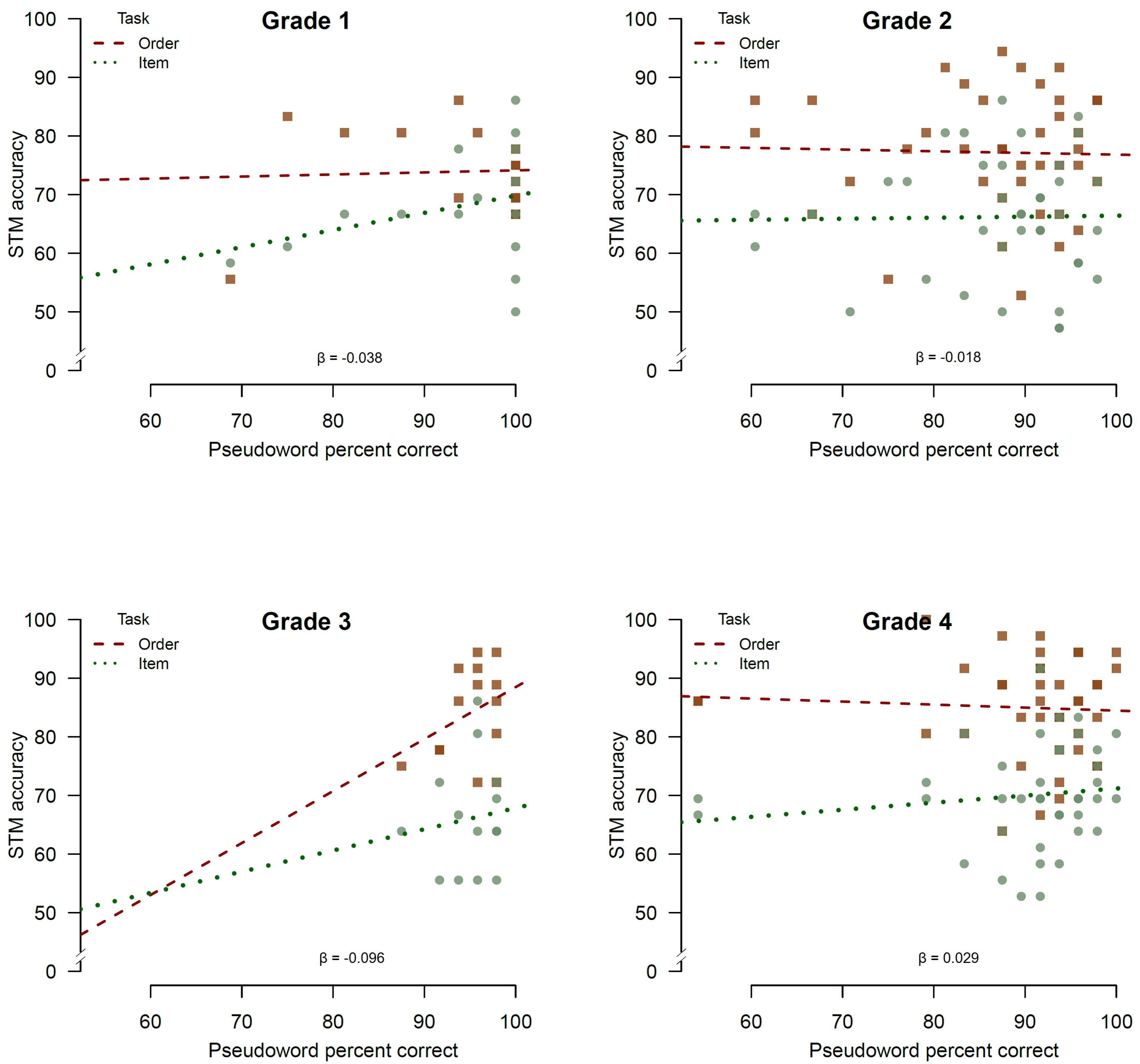


condition
item verbal
item nonverbal
order verbal
order nonverbal
100
95
90
85
80
75
70
65
60
55
0
percent correct
grade 1
grade 2
grade 3
grade 4

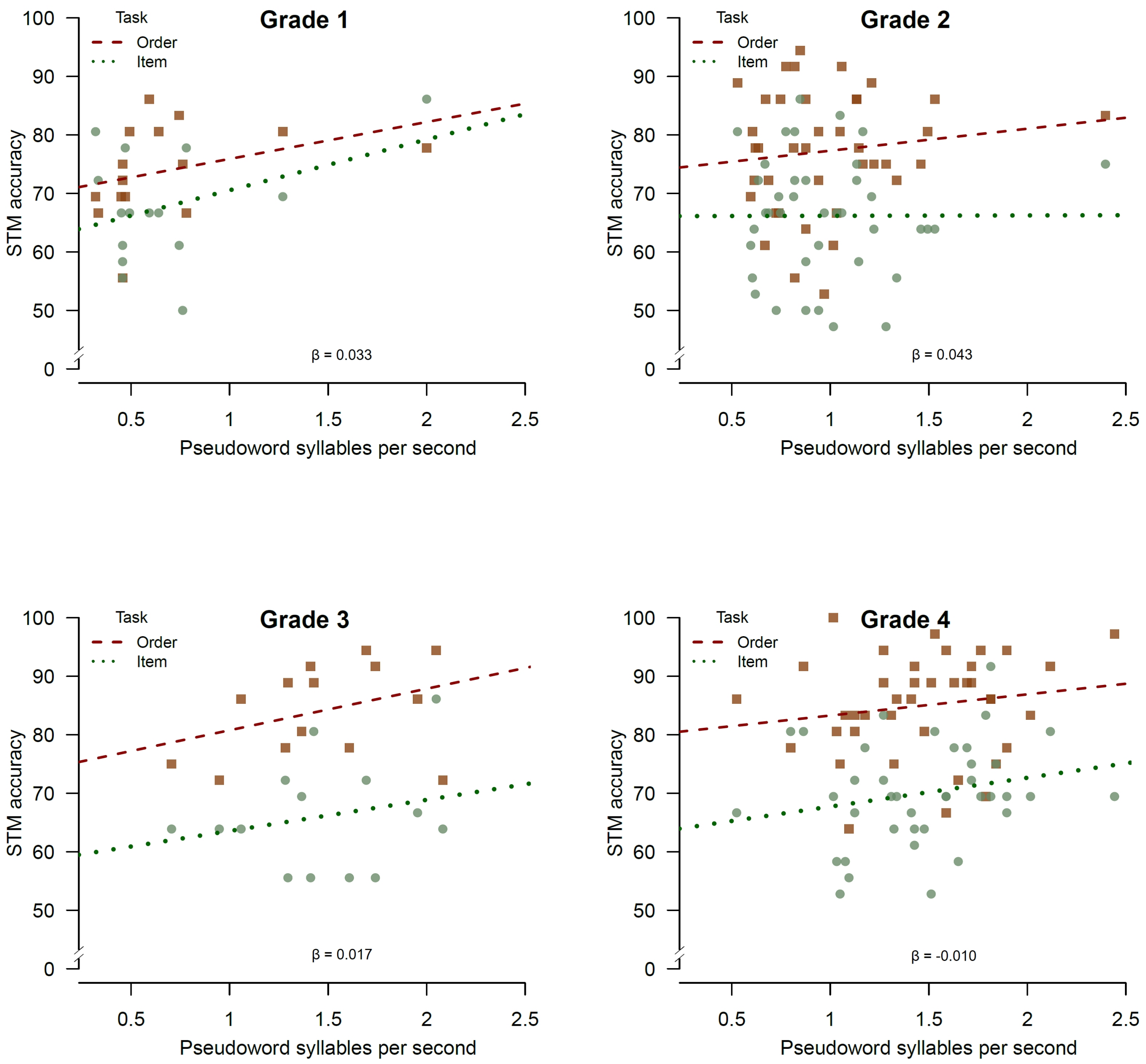


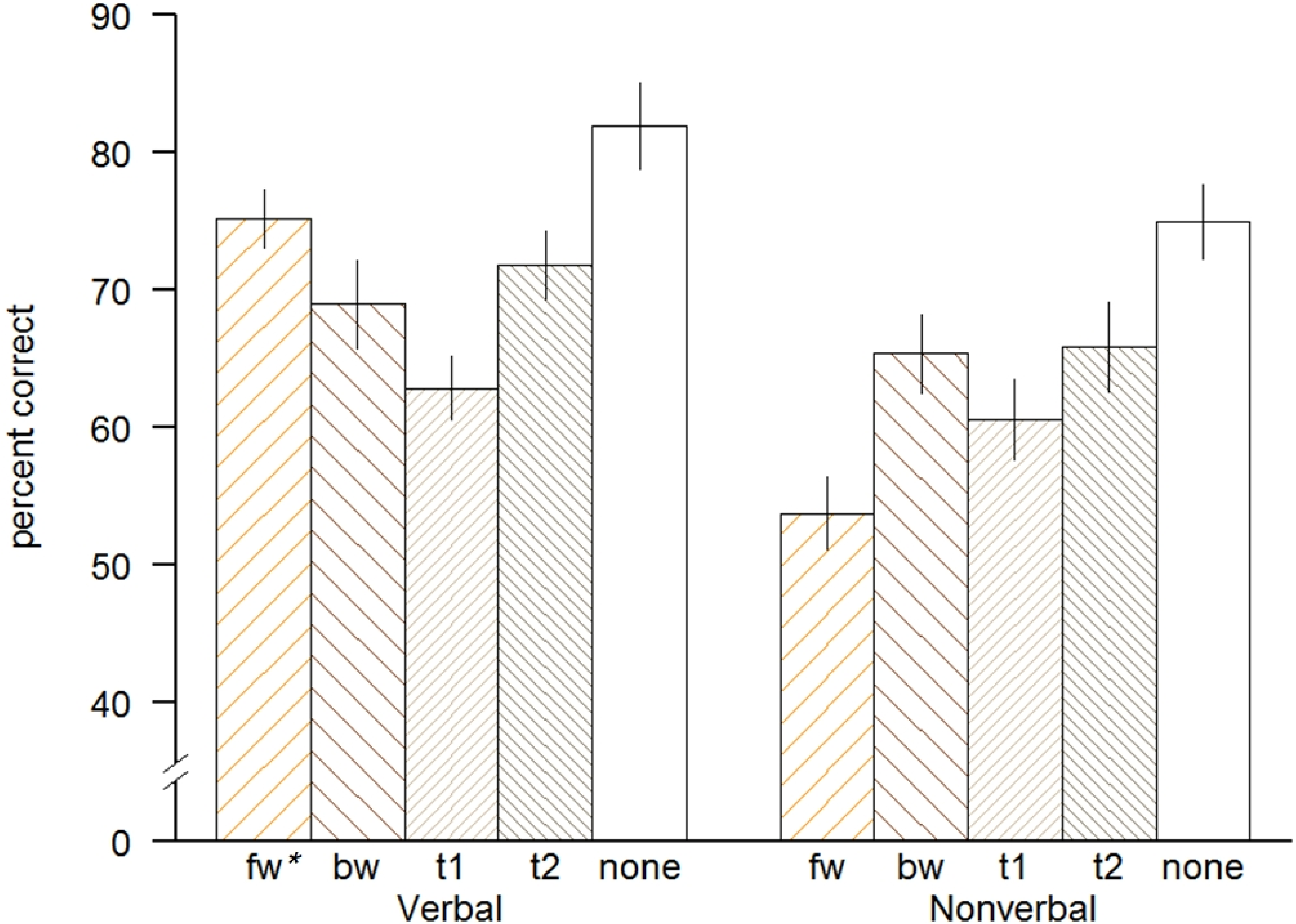
Grade 1
Grade 2
Grade 3
Grade 4
Task
Order
Item
STM accuracy
Word percent correct
β = -0.103
β = 0.061
β = -0.109
β = 0.052


Grade 1
Task
Order
Item
STM accuracy
Word syllables per second
β = 0.066
Grade 2
Task
Order
Item
STM accuracy
Word syllables per second
β = 0.042
Grade 3
Task
Order
Item
STM accuracy
Word syllables per second
β = 0.120
Grade 4
Task
Order
Item
STM accuracy
Word syllables per second
β = -0.023

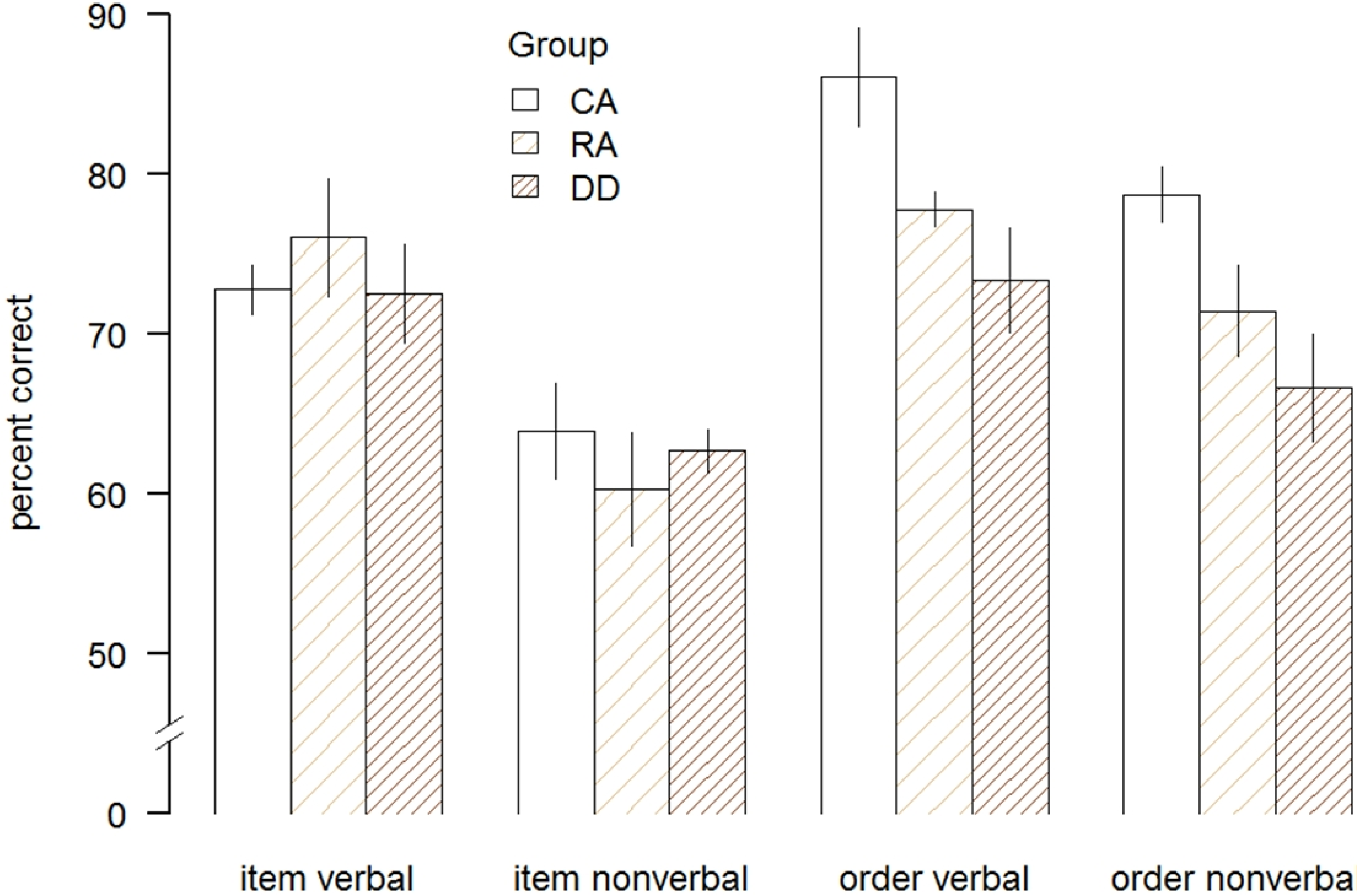


Grade 1
Grade 2
Grade 3
Grade 4
Task
Order
Item
STM accuracy
Pseudoword percent correct
β = -0.038
β = -0.018
β = -0.096
β = 0.029


Grade 1
Task
Order
Item
STM accuracy
Pseudoword syllables per second
β = 0.033
Grade 2
β = 0.043
Grade 3
β = 0.017
Grade 4
β = -0.010
100
90
80
70
60
50
0
0.5
1
1.5
2
2.5



90
80
70
60
50
40
0
percent correct
fw*
bw
t1
t2
none
Verbal
fw
bw
t1
t2
none
Nonverbal



Group
CA
RA
DD
90
80
70
60
50
0
percent correct
item verbal
item nonverbal
order verbal
order nonverbal

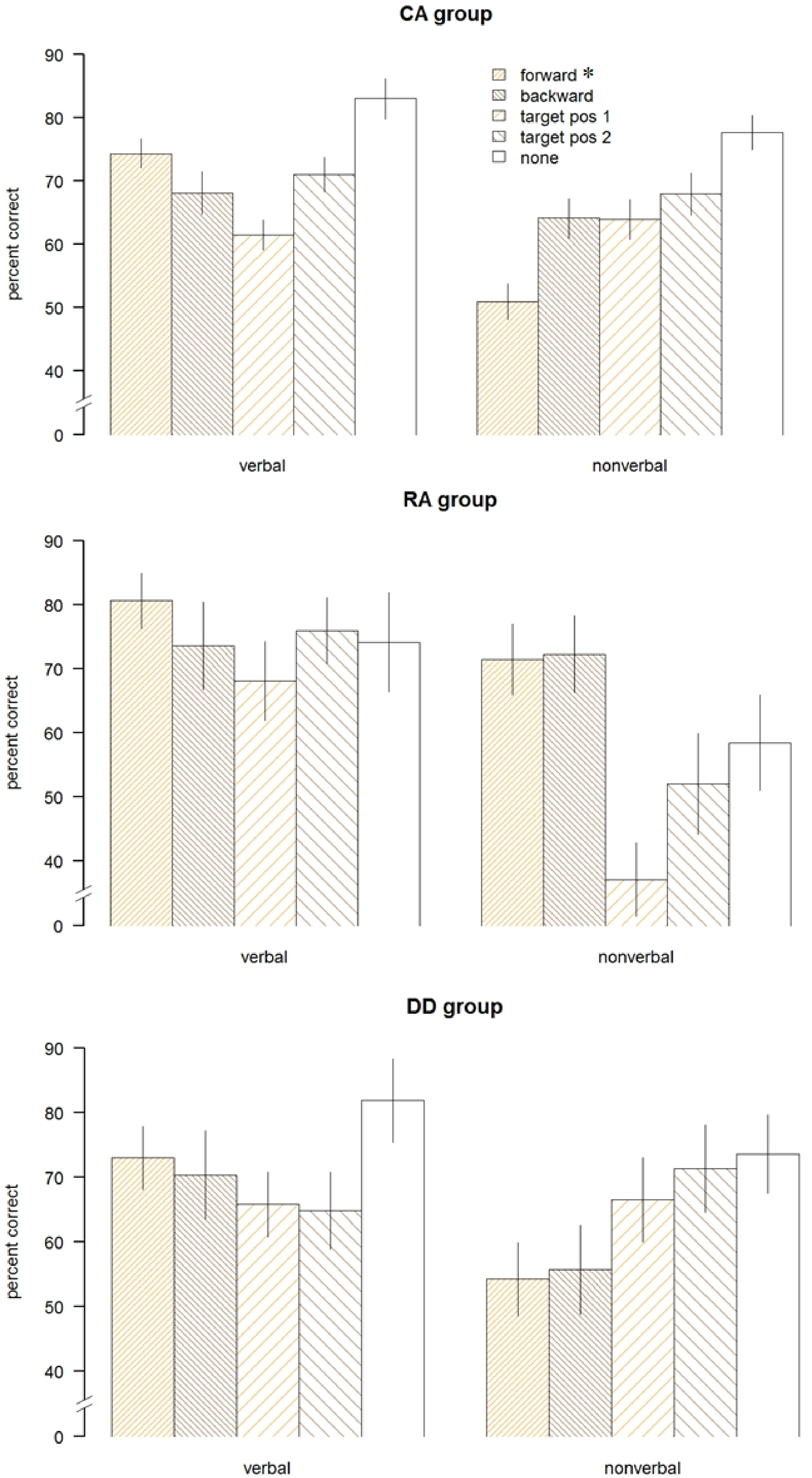

CA group
forward *
backward
target pos 1
target pos 2
none
percent correct
90
80
70
60
50
40
0
verbal
nonverbal
RA group
percent correct
90
80
70
60
50
40
0
verbal
nonverbal
DD group
percent correct
90
80
70
60
50
40
0
verbal
nonverbal

Table 1: Participant data of Experiment I by Grade

n = 107	*Grade 1*	*Grade 2*	*Grade 3*	*Grade 4*
Number of children	15	39	14	39
Bilinguals	1	7	2	15
Gender (female/male)	11 / 4	24 / 15	5 / 9	14 / 25
Age in months (mean(±SD))	82 (2.5)	92 (3.7)	106 (2.7)	112 (3.7)
Handedness (r/l/ambidex)	14 / 0 / 1	34 / 5 / 0	13 / 1 / 0	33 / 5 / 1

Table 2: Mean Word and Pseudoword Reading scores (±SD) by Grade

Measure	*Grade 1*	*Grade 2*		*Grade 3*		*Grade 4*	
		Word Reading					
Error perc.[n]	5.24 (5.2)	5.68 (5.12)		1.91 (1.80)	***	2.98 (3.29)	**
Speed	.81 (.66)	1.26 (.51)	***	2.26 (.65)	***	2.22 (.74)	***
Acc norm		.23 (.82)		.71 (.50)		- .11 (1.23)	
Speed norm		- .89 (1.26)		.14 (.89)		- .97 (1.53)	
		Pseudoword Reading					
Error perc.	6.94 (9.94)	13.41 (10.06)	***	4.91 (3.04)	***	9.67 (9.82)	
Speed	.68 (.42)	.98 (.35)	***	1.47 (.40)	***	1.47 (.39)	***
Acc norm		.11 (.97)		.73 (.29)		.09 (1.18)	
Speed norm		- .65 (1.00)		.16 (1.02)		- .67 (1.53)	

[n] Error perc. = percentage of mistakes to all words read; Speed = syllables per second; Acc norm = sample performance relative to norm for the number of mistakes per test (equal in sample and norm = 0), Speed norm = sample performance relative to norm for the overall time needed per test. Significant contrasts denote change with respect to all younger grades.

Significance codes also for the following tables and figures in this article: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' .1

Table 3: Mean number of correct responses[n] (±SD) of STM tasks by Grade

Task	*Domain*	*Grade 1*	*Grade 2*	*Grade 3*	*Grade 4*
Item	verbal	13.47 (1.92)	12.54 (2.33)	13.14 (2.32)	13.08 (1.92)
	nonverbal	10.93 (2.49)	11.28 (2.18)	10.64 (2.59)	12.13 (2.21)
Order	verbal	13.87 (2.56)	14.56 (2.28)	15.79 (2.12)	15.97 (1.78)
	nonverbal	12.73 (1.87)	13.23 (2.16)	14.50 (1.74)	14.62 (2.11)

[n] Correct trials out of a possible maximum of 18

Table 4: Percentage of correct responses in item tasks for Chronology by Domain and Grade

	Domain	*forward**	*backward*	*target 1*	*target 2*	*none*
	verbal	74.58	68.69	63.24	71.34	81.00
	nonverbal	54.21	65.05	60.44	66.67	74.14
	Mean	65.52	66.67	61.84	69.00	77.57
			Grade			
1	verbal	80.27	73.45	68.38	75.88	74.16
	nonverbal	71.43	72.23	37.53	51.99	58.30
2	verbal	70.18	64.10	62.65	68.64	85.22
	nonverbal	48.79	66.33	65.87	62.95	70.70
3	verbal	81.83	64.26	68.53	69.27	79.39
	nonverbal	44.05	54.32	61.89	60.69	81.40
4	verbal	75.12	73.26	58.30	73.86	81.93
	nonverbal	55.26	65.26	63.25	75.20	82.52

* *forward*: both probes were in the list in same order; *backward*: both probes were in the list in opposite order; *target 1*: only first probe was in the list; *target 2*: only second probe was in the list; *none*: neither probe was in the list

Table 5: Participant data of Experiment II by Group

n = 123	*DD*	*RA*[n]	*CA*
Number of children	16	15	92
Bilinguals	1	1	24
Gender (female/male)	7 / 9	11 / 4	43 / 49
Age in months (mean(SD))	99.4 (13)	82 (2.5)	102.7 (10.1)
Handedness (r/l/ambidex)	14 / 1 / 1	14 / 0 / 1	80 / 11 / 1
block design test, WISC-III	11.8 (2.7)	14.2 (2.4)	13.4 (2.9)
Word reading error percent	32.64 (20.93)	5.24 (5.2)	3.96 (4.29)
Word reading speed (syll/sec)	.83 (.54)	.81 (.66)	1.82 (.8)
Pseudoword reading err perc.	42.48 (14.6)	6.94 (9.94)	10.53 (9.66)
Pseudoword reading speed	.7 (.41)	.68 (.42)	1.26 (.45)

[n] Reading-age matched control group *RA* = grade 1

Table 6: Mean number of correct responses[n] (±SD) of STM tasks by Group

Task	*Domain*	*DD*	*RA*	*CA*
Item	verbal	12.81 (1.38)	13.47 (1.92)	12.86 (2.16)
	nonverbal	11.25 (2.38)	10.93 (2.49)	11.54 (2.30)
Order	verbal	13.06 (2.26)	13.87 (2.56)	15.35 (2.15)
	nonverbal	11.88 (3.56)	12.73 (1.87)	14.01 (2.17)

[n] Correct trials out of a possible maximum of 18

Appendix

Material table for verbal short-term memory tasks

Word	*Type Freq*	*% acc*^{n}	*H*	*Fam Mean*	*Fam SD*	*AoA Mean*	*AoA SD*
cane	2.78	100	0	4.20	1.2	2.05	1.26
casa	3.93	92.5	0.43	3.50	1.5	1.92	1.02
casco	2.06	96.86	0.16	3.44	1.2	3.33	1.41
chiave	2.28	100	0	4.60	1.5	3.80	1.38
chiodo	1.41	100	0	3.27	1.3	4.30	1.56
cigno	1.70	92.5	0.48	2.22	1.3	4.62	1.74
cuore	2.80	100	0	4.05	1.3	3.35	1.80
fiocco	1.08	97.5	0.16	3.22	1.6	3.80	1.42
fiore	1.97	100	0	3.82	1.2	2.17	1.13
foca	0.60	95	0.27	1.65	1.1	4.80	1.71
fungo	2.01	100	0	2.95	1.5	4.10	1.46
gallo	2.29	85	0.57	2.97	1.3	2.95	1.32
gatto	2.57	97.5	0.16	3.85	1.4	2.45	1.08
gonna	2.17	95	0.32	3.60	1.3	3.00	1.20
guanto	1.75	100	0	3.42	1.5	3.78	1.56
letto	3.00	90	0.48	4.77	0.6	2.15	1.16
libro	2.80	100	0	4.77	0.6	3.10	1.55
mano	2.99	95	0.32	4.82	0.4	1.90	1.15
mela	2.26	100	0	4.32	1.1	2.15	1.12
mosca	2.01	92.5	0.31	3.17	1.3	3.33	1.47
mucca	1.69	92.5	0.35	2.70	1.3	3.10	1.43
naso	2.10	100	0	4.47	0.8	2.03	1.22
occhio	2.51	100	0	4.75	0.5	1.75	1.10
orso	2.35	95	0.32	1.82	1.2	3.88	1.64
pera	1.74	100	0	4.12	1.1	2.50	1.20

pesce	2.26	95	0.32	3.60	1.4	3.10	1.35
Word	*Type Freq*	*% acc*[n]	*H*	*Fam Mean*	*Fam SD*	*AoA Mean*	*AoA SD*
piede	2.20	95	0.27	4.70	0.7	2.47	1.49
pipa	1.83	100	0	2.45	1.7	4.60	1.41
porta	2.74	97.5	0.16	4.50	0.9	3.05	1.55
pozzo	2.17	100	0	1.97	1.3	4.97	1.41
ragno	1.26	100	0	2.17	1.4	3.42	1.41
rana	1.00	97.5	0.16	2.27	1.2	3.62	1.53
ruota	1.83	97.5	0.16	2.37	1.0	3.78	1.31
scala	2.05	87.5	0.6	3.52	1.3	4.10	1.77
scarpa	1.97	97.5	0.16	4.65	1.4	2.88	1.42
scopa	1.39	95	0.27	4.00	0.8	3.38	1.29
sedia	2.40	100	0	4.57	0.9	2.65	1.29
sega	0.70	100	0	2.40	1.5	5.30	1.57
sole	2.67	100	0	4.77	0.4	1.70	0.79
stella	1.94	100	0	4.17	1.2	3.00	1.22
tigre	2.17	97.5	0.16	2.07	1.5	4.32	1.72
topo	2.58	100	0	2.22	1.6	3.10	1.50
torta	1.73	100	0	4.20	0.9	2.52	0.96
treno	2.77	92.5	0.35	4.22	1.2	3.37	1.88
tromba	1.87	87.5	0.41	2.30	1.4	4.97	1.94
vaso	1.74	95	0.32	3.12	1.4	4.07	1.61
zebra	1.20	100	0	1.92	1.4	4.55	1.48
zucca	1.41	97.5	0	2.47	1.2	4.70	1.68
Mean	2.06	96.86	0.16	3.44	1.18	3.33	1.41
SD	0.66	3.92	0.18		0.33		0.25
Min	0.60	85.00	0.00	1.65	0.40	1.70	0.79
Max	3.93	100.00	0.60	4.82	1.70	5.30	1.94

[n] Type Freq = log of type frequency; % acc = percentage of accuracy of picture naming; H = degree of naming agreement, 0=full agreement; Fam = familiarity, daily contact on a scale from 1 to 5; AoA = age of acquisition, rating of word learning age.

Regression tables Experiment I

Linear Regressions of reading measures by Grade					
	Estimate	*Std. Error*	*t value*	*p*	
Word reading error percentage					
Intercept	3.951	0.229	17.238	< 2e-16	***
Grade 2	0.220	0. 320	0.687	0.4927	
Grade 3	-1.182	0.216	-5.472	7.61e-08	***
Grade 4	-0.325	0.110	-2.947	0.0034	**
Word reading speed					
Intercept	1.639	0.035	46.907	< 2e-16	***
Grade 2	0.224	0.049	4.589	5.87e-06	***
Grade 3	0.409	0.033	12.419	< 2e-16	***
Grade 4	0.194	0.017	11.555	< 2e-16	***
Pseudoword reading error percentage					
Intercept	8.733	0.509	17.170	< 2e-16	***
Grade 2	3.232	0.710	4.551	6.99e-06	***
Grade 3	-1.755	0.479	-3.664	0.0003	***
Grade 4	0.312	0.245	1.275	0.2031	
Pseudoword reading speed					
Intercept	1.149	0.021	55.212	< 2e-16	***
Grade 2	0.147	0.029	5.057	6.34e-07	***
Grade 3	0.215	0.020	10.953	< 2e-16	***
Grade 4	0.105	0.010	10.530	< 2e-16	***

Omnibus ANOVA on linear mixed model of STM by Grade				
	Chisquare	*Df*	*p*	
Intercept	302.129	1	< 2.2e-16	***
Task	123.696	1	< 2.2e-16	***
Domain	67.341	1	2.3e-16	***
Grade	19.078	3	0.0003	***
Task : Domain	0.232	1	0.6299	
Task : Grade	21.222	3	9.5e-05	***
Domain : Grade	1.700	3	0.6369	
Task : Domain : Grade	4.566	3	0.2064	

ANOVA on linear mixed models of STM by Grade and reading				
	Chisquare	*Df*	*p*	
Word reading error percentage				
Intercept	229.388	1	< 2.2e-16	***
Task	66.674	1	3.2e-16	***
Domain	47.096	1	6.8e-12	***
Grade	13.165	3	0.0043	**
Word err perc	2.394	1	0.1218	
Task : Domain	0.044	1	0.8330	
Task : Grade	13.679	3	0.0034	**
Domain : Grade	1.601	3	0.6592	
Task : Word err perc	3.809	1	0.0510	.
Domain : Word err perc	0.301	1	0.5831	

Grade : Word err perc	0.656	3	0.8834	
Task: Domain : Grade	3.379	3	0.3368	
Task : Domain : Word err perc	0.696	1	0.4042	
Task : Grade : Word err perc	7.861	3	0.0490	*
Domain : Grade : Word err perc	0.740	3	0.8640	
Task : Domain : Grade : Word err perc	3.311	3	0.3461	
Word reading speed				
Intercept	196.486	1	< 2.2e-16	***
Task	40.542	1	1.9e-10	***
Domain	38.819	1	4.7e-10	***
Grade	5.952	3	0.1140	
Word syll sec	7.292	1	0.0070	**
Task : Domain	0.004	1	0.9521	
Task : Grade	9.759	3	0.0207	*
Domain : Grade	0.966	3	0.8094	
Task : Word syll sec	2.192	1	0.1387	
Domain : Word syll sec	1.288	1	0.2565	
Grade : Word syll sec	3.478	3	0.3236	
Task: Domain : Grade	4.413	3	0.2202	
Task : Domain : Word syll sec	1.787	1	0.1813	
Task : Grade : Word syll sec	8.625	3	0.0347	*
Domain : Grade : Word syll sec	3.294	3	0.3484	
Task : Domain : Grade : Word syll sec	4.837	3	0.1841	
Pseudoword reading error percentage				
Intercept	179.976	1	< 2.2e-16	***
Task	54.435	1	1.6e-13	***
Domain	31.702	1	1.8e-08	***
Grade	18.808	3	0.0003	***
Pword err perc	1.638	1	0.2006	
Task : Domain	1.483	1	0.2233	
Task : Grade	14.590	3	0.0022	**
Domain : Grade	0.471	3	0.9252	
Task : Pword err perc	0.337	1	0.5617	
Domain : Pword err perc	0.525	1	0.4688	
Grade : Pword err perc	1.589	3	0.6619	
Task: Domain : Grade	7.146	3	0.0674	.
Task : Domain : Pword err perc	1.399	1	0.2370	
Task : Grade : Pword err perc	1.838	3	0.6066	
Domain : Grade : Pword err perc	1.046	3	0.7900	
Task : Domain : Grade : Pword err perc	6.264	3	0.0995	.
Pseudoword reading speed				
Intercept	226.156	1	< 2.2e-16	***
Task	65.392	1	6.1e-16	***
Domain	33.408	1	7.5e-09	***
Grade	5.442	3	0.1421	
Pword syll sec	7.646	1	0.0057	**
Task : Domain	0.159	1	0.6901	
Task : Grade	9.019	3	0.0290	*
Domain : Grade	0.132	3	0.9877	
Task : Pword syll sec	0.718	1	0.3969	
Domain : Pword syll sec	1.421	1	0.2333	

Grade : Pword syll sec	1.220	3	0.7482
Task: Domain : Grade	2.694	3	0.4412
Task : Domain : Pword syll sec	0.076	1	0.7830
Task : Grade : Pword syll sec	1.106	3	0.7756
Domain : Grade : Pword syll sec	1.660	3	0.6459
Task : Domain : Grade : Pword syll sec	3.730	3	0.2921

Omnibus ANOVA on linear mixed model of Chronology by Domain

	Chisquare	*Df*	*p*	
Intercept	216.301	1	< 2.2e-16	***
Chronology	37.029	4	1.8e-07	***
Domain	25.070	1	5.6e-07	***
Chronology : Domain	15.989	4	0.0030	**

ANOVA on linear mixed model of Chronology by Domain and Grade

	Chisquare	*Df*	*p*	
Intercept	175.225	1	< 2.2e-16	***
Chronology	25.937	4	3.3e-05	***
Domain	27.720	1	1.4e-07	***
Grade	3.699	3	0.2958	
Chronology : Domain	11.434	4	0.0221	*
Chronology : Grade	29.990	12	0.0028	**
Domain : Grade	6.124	3	0.1057	
Chronology : Domain : Grade	21.214	12	0.0473	*

Regression tables Experiment II

Omnibus ANOVA on linear mixed model of STM by Group				
	Chisquare	*Df*	*p*	
Intercept	350.745	1	< 2.2e-16	***
Task	161.093	1	< 2.2e-16	***
Domain	58.218	1	2.4e-14	***
Group	13.297	2	0.00130	**
Task : Domain	0.662	1	0.4159	
Task : Group	28.899	2	5.3e-07	***
Domain : Group	0.649	2	0.7230	
Task : Domain : Group	3.042	2	0.2185	

ANOVA on linear mixed models of STM by reading, per Group				
	Chisquare	*Df*	*p*	
Word reading error percentage				
	Group CA			
Intercept	317.856	1	< 2.2e-16	***
Task	161.620	1	< 2.2e-16	***
Domain	58.404	1	2.1e-14	***
Word err perc	5.999	1	0.0143	*
Task : Domain	0.809	1	0.3684	
Task : Word err perc	0.529	1	0.4668	
Domain : Word err perc	1.424	1	0.2327	
Task : Domain : Word err perc	0.206	1	0.6497	
	Group RA			
Intercept	97.016	1	< 2.2e-16	***
Task	4.302	1	0.0381	*
Domain	15.392	1	8.7e-05	***
Word err perc	0.966	1	0.3258	
Task : Domain	1.476	1	0.2245	
Task : Word err perc	3.617	1	0.0572	.
Domain : Word err perc	1.666	1	0.1967	
Task : Domain : Word err perc	4.140	1	0.0419	*
	Group DD			
Intercept	77.861	1	< 2.2e-16	***
Task	0.616	1	0.4325	
Domain	7.324	1	0.0068	**
Word err perc	0.680	1	0.4095	
Task : Domain	0.043	1	0.8352	
Task : Word err perc	0.000	1	0.9930	
Domain : Word err perc	0.170	1	0.6802	
Task : Domain : Word err perc	3.185	1	0.0743	.
Word reading speed				
	Group CA			
Intercept	329.398	1	< 2.2e-16	***
Task	165.112	1	< 2.2e-16	***
Domain	58.185	1	2.4e-14	***
Word syll sec	14.223	1	0.0002	***
Task : Domain	0.961	1	0.3269	

Task : Word syll sec	8.627	1	0.0033	**
Domain : Word syll sec	0.418	1	0.5178	
Task : Domain : Word syll sec	0.496	1	0.4813	
Group RA				
Intercept	100.425	1	< 2.2e-16	***
Task	3.638	1	0.0565	.
Domain	14.688	1	0.0001	***
Word syll sec	2.964	1	0.0852	.
Task : Domain	1.877	1	0.1707	
Task : Word syll sec	0.482	1	0.4877	
Domain : Word syll sec	0.809	1	0.3685	
Task : Domain : Word syll sec	0.198	1	0.6567	
Group DD				
Intercept	114.154	1	< 2.2e-16	***
Task	0.771	1	0.3800	
Domain	3.634	1	0.0566	.
Word syll sec	9.754	1	0.0018	**
Task : Domain	0.522	1	0.4700	
Task : Word syll sec	2.963	1	0.0852	.
Domain : Word syll sec	0.092	1	0.7623	
Task : Domain : Word syll sec	1.974	1	0.1600	
Pseudoword reading error percentage				
Group CA				
Intercept	310.092	1	< 2.2e-16	***
Task	160.667	1	< 2.2e-16	***
Domain	58.056	1	2.6e-14	***
Pword err perc	0.967	1	0.3254	
Task : Domain	0.706	1	0.4007	
Task : Pword err perc	0.049	1	0.8245	
Domain : Pword err perc	3.320	1	0.0684	.
Task : Domain : Pword err perc	0.608	1	0.4354	
Group RA				
Intercept	95.389	1	< 2.2e-16	***
Task	3.762	1	0.0524	.
Domain	14.244	1	0.0002	***
Pword err perc	0.870	1	0.3510	
Task : Domain	1.818	1	0.1776	
Task : Pword err perc	0.486	1	0.4860	
Domain : Pword err perc	0.110	1	0.7408	
Task : Domain : Pword err perc	3.995	1	0.0456	*
Group DD				
Intercept	71.228	1	< 2e-16	***
Task	0.368	1	0.5440	
Domain	6.607	1	0.0102	*
Pword err perc	0.615	1	0.4328	
Task : Domain	0.000	1	0.9896	
Task : Pword err perc	1.034	1	0.3093	
Domain : Pword err perc	< 0.001	1	0.9904	
Task : Domain : Pword err perc	0.475	1	0.4908	
Pseudoword reading speed				
Group CA				

Intercept	327.656	1	< 2.2e-16	***
Task	164.489	1	< 2.2e-16	***
Domain	58.955	1	1.6e-14	***
Pword syll sec	13.194	1	0.0003	***
Task : Domain	0.980	1	0.3222	
Task : Pword syll sec	7.189	1	0.0073	**
Domain : Pword syll sec	2.039	1	0.1533	
Task : Domain : Pword syll sec	0.190	1	0.6628	
Group RA				
Intercept	109.257	1	< 2.2e-16	***
Task	3.736	1	0.0533	.
Domain	14.322	1	0.0002	***
Pword syll sec	5.116	1	0.0237	*
Task : Domain	1.819	1	0.1774	
Task : Pword syll sec	0.038	1	0.8463	
Domain : Pword syll sec	0.005	1	0.9459	
Task : Domain : Pword syll sec	0.118	1	0.7308	
Group DD				
Intercept	65.886	1	4.8e-16	***
Task	0.261	1	0.6096	
Domain	6.296	1	0.0121	*
Pword syll sec	0.159	1	0.6899	
Task : Domain	0.001	1	0.9822	
Task : Pword syll sec	0.007	1	0.9360	
Domain : Pword syll sec	0.922	1	0.3369	
Task : Domain : Pword syll sec	4.936	1	0.0263	*

ANOVA on linear mixed model of Chronology by Domain and Group

	Chisquare	*Df*	*p*	
Intercept	204.491	1	< 2.2e-16	***
Chronology	44.664	4	4.7e-09	***
Domain	14.555	1	0.0001	***
Group	0.383	2	0.8258	
Chronology : Domain	20.338	4	0.0004	***
Chronology : Group	23.486	8	0.0028	**
Domain : Group	3.992	2	0.1358	
Chronology : Domain : Group	14.264	8	0.0751	.

ANOVA on linear mixed models of Chronology by Domain, reading, and Group

	Chisquare	*Df*	*p*	
Word reading error percentage				
Intercept	111.521	1	< 2.2e-16	***
Chronology	34.830	4	5.0e-07	***
Domain	7.476	1	0.0063	**
Group	0.866	2	0.6484	
Word err perc	1.678	1	0.1952	
Chronology : Domain	10.244	4	0.0365	*
Chronology : Group	12.037	8	0.1496	
Domain : Group	5.020	2	0.0813	.
Chronology : Word err perc	9.451	4	0.0508	.

Domain : Word err perc	0.074	1	0.7848	
Group : Word err perc	2.447	2	0.2942	
Chronology : Domain : Group	11.874	8	0.1569	
Chronology : Domain : Word err perc	2.671	4	0.6144	
Chronology : Group : Word err perc	24.258	8	0.0021	**
Domain : Group : Word err perc	0.768	2	0.6811	
Chron : Domain : Group : W. err perc	5.048	8	0.7524	
Word reading speed				
Intercept	188.782	1	< 2.2e-16	***
Chronology	40.488	4	3.4e-08	***
Domain	14.330	1	0.0002	***
Group	3.075	2	0.2149	
Word syll sec	2.749	1	0.0973	.
Chronology : Domain	17.593	4	0.0015	**
Chronology : Group	13.318	8	0.1014	
Domain : Group	1.895	2	0.3877	
Chronology : Word syll sec	10.705	4	0.0301	*
Domain : Word syll sec	0.314	1	0.5755	
Group : Word syll sec	2.071	2	0.3551	
Chronology : Domain : Group	4.895	8	0.7687	
Chronology : Domain : Word syll sec	8.240	4	0.0832	.
Chronology : Group : Word syll sec	5.470	8	0.7063	
Domain : Group : Word syll sec	0.490	2	0.7827	
Chron : Domain : Group : W. syll sec	4.952	8	0.7627	
Pseudoword reading error percentage				
Intercept	180.855	1	< 2.2e-16	***
Chronology	40.615	4	3.2e-08	***
Domain	9.417	1	0.0021	**
Group	0.135	2	0.9346	
Pword err perc	0.427	1	0.5133	
Chronology : Domain	16.880	4	0.0020	**
Chronology : Group	8.968	8	0.3450	
Domain : Group	8.100	2	0.0174	*
Chronology : Pword err perc	3.311	4	0.5072	
Domain : Pword err perc	3.028	1	0.0819	.
Group : Pword err perc	0.051	2	0.9749	
Chronology : Domain : Group	10.560	8	0.2279	
Chronology : Domain : Pword err perc	1.459	4	0.8338	
Chronology : Group : Pword err perc	23.501	8	0.0028	**
Domain : Group : Pword err perc	6.788	2	0.0336	*
Chron : Domain : Group : Pw. err perc	2.447	8	0.9642	
Pseudoword reading speed				
Intercept	177.702	1	< 2.2e-16	***
Chronology	37.759	4	1.3e-07	***
Domain	12.310	1	0.0005	***
Group	2.178	2	0.3365	
Pword syll sec	3.582	1	0.0584	.
Chronology : Domain	18.596	4	0.0009	***
Chronology : Group	12.504	8	0.1301	
Domain : Group	0.645	2	0.7244	
Chronology : Pword syll sec	8.079	4	0.0887	.

Domain : Pword syll sec	0.182	1	0.6701
Group : Pword syll sec	2.104	2	0.3492
Chronology : Domain : Group	8.586	8	0.3784
Chronology : Domain : Pword syll sec	4.168	4	0.3837
Chronology : Group : Pword syll sec	6.321	8	0.6113
Domain : Group : Pword syll sec	1.592	2	0.4511
Chron : Domain : Group : Pw. syll sec	5.452	8	0.7083

Supplementary Materials

Control Models for Bilingualism

Omnibus ANOVA on linear mixed model of STM by Group with **Bilingualism**				
	Chisquare	*Df*	*p*	
Intercept	278.676	1	< 2.2e-16	***
Task	112.117	1	< 2.2e-16	***
Domain	54.386	1	1.6e-13	***
Grade	17.910	3	0.0005	***
Bilingualism (Bil)	0.125	1	0.7237	
Task : Domain	0.064	1	0.8008	
Task : Grade	15.767	3	0.0013	**
Domain : Grade	0.849	3	0.8377	
Task : Bilingualism	3.099	1	0.0784	.
Domain : Bilingualism	0.001	1	0.9772	
Grade : Bilingualism	1.846	3	0.6049	
Task : Domain : Grade	1.285	3	0.7327	
Task : Domain : Bil	2.386	1	0.1224	
Task : Grade : Bil	3.897	3	0.2728	
Domain : Grade : Bil	3.106	3	0.3755	
Task : Domain : Grade : Bil	4.536	3	0.2092	

ANOVA on linear mixed models of STM by Grade and reading with **Bilingualism**				
	Chisquare	*Df*	*p*	
Word reading error percentage				
Intercept	180.597	1	< 2.2e-16	***
Task	57.956	1	2.7e-14	***
Domain	22.000	1	2.7e-06	***
Grade	8.704	3	0.0335	*
Word err perc	1.104	1	0.2933	
Bil	0.002	1	0.9625	
Task : Domain	0.321	1	0.5709	
Task : Grade	8.789	3	0.0322	*
Domain : Grade	3.304	3	0.3471	
Task : Word err perc	0.438	1	0.5081	
Domain : Word err perc	2.616	1	0.1058	
Grade : Word err perc	1.510	3	0.6801	
Task : Bil	0.001	1	0.9715	
Domain : Bil	0.001	1	0.9712	
Grade : Bil	2.240	3	0.5241	
Word err perc : Bil	0.573	1	0.4489	
Task : Domain : Grade	0.734	3	0.8652	
Task : Domain : Word err perc	2.044	1	0.1528	
Task : Grade : Word err perc	8.414	3	0.0382	*
Domain : Grade : Word err perc	2.347	3	0.5036	
Task : Domain : Bil	0.002	1	0.9694	
Task : Grade : Bil	0.972	3	0.8080	

Domain : Grade : Bil	3.795	3	0.2845	
Task : Word err perc : Bil	1.637	1	0.2007	
Domain : Word err perc : Bil	1.985	1	0.1589	
Grade : Word err perc : Bil	0.439	2	0.8028	
Task : Domain : Grade : Word err perc	1.722	3	0.6321	
Task : Domain : Grade : Bil	0.989	3	0.8039	
Task : Domain : Word err perc : Bil	1.753	1	0.1855	
Task : Grade : Word err perc : Bil	0.002	2	0.9992	
Domain : Grade : Word err perc : Bil	0.786	2	0.6752	
Task : Domain : Grade: Word err perc : Bil	1.231	2	0.5405	
Word reading speed				
Intercept	154.112	1	< 2.2e-16	***
Task	33.503	1	7.1e-09	***
Domain	25.449	1	4.5e-07	***
Grade	6.933	3	0.0741	.
Word syll sec	2.765	1	0.0963	.
Bil	0.002	1	0.9677	
Task : Domain	0.006	1	0.9396	
Task : Grade	7.592	3	0.0552	.
Domain : Grade	1.206	3	0.7515	
Task : Word syll sec	0.355	1	0.5513	
Domain : Word syll sec	0.860	1	0.3537	
Grade : Word syll sec	6.409	3	0.0933	.
Task : Bil	0.010	1	0.9766	
Domain : Bil	0.002	1	0.9671	
Grade : Bil	2.401	3	0.4935	
Word err perc : Bil	0.224	1	0.6363	
Task : Domain : Grade	1.610	3	0.6571	
Task : Domain : Word syll sec	0.391	1	0.5319	
Task : Grade : Word syll sec	7.791	3	0.0505	.
Domain : Grade : Word syll sec	2.132	3	0.5455	
Task : Domain : Bil	0.002	1	0.9612	
Task : Grade : Bil	0.007	3	0.9999	
Domain : Grade : Bil	0.233	3	0.9721	
Task : Word syll sec : Bil	3.128	1	0.0770	
Domain : Word syll sec : Bil	0.443	1	0.5057	.
Grade : Word syll sec : Bil	0.677	2	0.7129	
Task : Domain : Grade : Word syll sec	1.640	3	0.6504	
Task : Domain : Grade : Bil	2.264	3	0.5194	
Task : Domain : Word syll sec : Bil	0.011	1	0.9173	
Task : Grade : Word syll sec : Bil	0.273	2	0.8724	
Domain : Grade : Word syll sec : Bil	1.804	2	0.4058	
Task : Domain : Grade: Word syll sec : Bil	0.479	2	0.7871	
Pseudoword reading error percentage				
Intercept	171.837	1	< 2.2e-16	***
Task	54.016	1	2e-13	***
Domain	26.466	1	2.7e-07	***
Grade	18.996	3	0.0003	***
Pword err perc	0.792	1	0.3736	

Bil	0.004	1	0.9478	
Task : Domain	1.844	1	0.1745	
Task : Grade	12.342	3	0.0063	**
Domain : Grade	1.173	3	0.7595	
Task : Pword err perc	0.027	1	0.8704	
Domain : Pword err perc	0.866	1	0.3522	
Grade : Pword err perc	4.490	3	0.2132	
Task : Bil	0.003	1	0.9534	
Domain : Bil	0.001	1	0.9706	
Grade : Bil	6.025	3	0.1104	
Pword err perc : Bil	1.883	1	0.1700	
Task : Domain : Grade	3.175	3	0.3654	
Task : Domain : Pword err perc	3.527	1	0.0604	.
Task : Grade : Pword err perc	7.765	3	0.0511	.
Domain : Grade : Pword err perc	1.314	3	0.7259	
Task : Domain : Bil	0.003	1	0.9581	
Task : Grade : Bil	6.384	3	0.0943	.
Domain : Grade : Bil	1.053	3	0.7884	
Task : Pword err perc : Bil	4.870	1	0.0273	*
Domain : Pword err perc : Bil	0.443	1	0.5056	
Grade : Pword err perc : Bil	1.944	2	0.3783	
Task : Domain : Grade : Pword err perc	5.334	3	0.1489	
Task : Domain : Grade : Bil	3.592	3	0.3090	
Task : Domain : Pword err perc : Bil	0.477	1	0.4900	
Task : Grade : Pword err perc : Bil	0.231	2	0.8907	
Domain : Grade : Pword err perc : Bil	0.193	2	0.9080	
Task : Domain : Grade: Pword err perc : Bil	0.040	2	0.9801	
Pseudoword reading speed				
Intercept	193.745	1	< 2.2e-16	***
Task	57.993	1	2.6e-14	***
Domain	23.708	1	1.1e-06	***
Grade	6.415	3	0.0931	.
Pword syll sec	6.026	1	0.0141	*
Bil	0.003	1	0.9599	
Task : Domain	0.003	1	0.9566	
Task : Grade	8.800	3	0.0321	*
Domain : Grade	0.248	3	0.9694	
Task : Pword syll sec	0.509	1	0.4757	
Domain : Pword syll sec	0.962	1	0.3266	
Grade : Pword syll sec	4.903	3	0.1790	
Task : Bil	0.002	1	0.9692	
Domain : Bil	0.001	1	0.9722	
Grade : Bil	2.364	3	0.5004	
Pword syll sec : Bil	0.351	1	0.5538	
Task : Domain : Grade	1.284	3	0.7330	
Task : Domain : Pword syll sec	0.287	1	0.5923	
Task : Grade : Pword syll sec	4.650	3	0.1993	
Domain : Grade : Pword syll sec	1.878	3	0.5982	
Task : Domain : Bil	0.001	1	0.9719	
Task : Grade : Bil	3.091	3	0.3778	

Domain : Grade : Bil	1.602	3	0.6589	
Task : Pword syll sec : Bil	3.448	1	0.0633	.
Domain : Pword syll sec : Bil	0.360	1	0.5487	
Grade : Pword syll sec : Bil	0.048	2	0.9761	
Task : Domain : Grade : Pword syll sec	1.117	3	0.7730	
Task : Domain : Grade : Bil	0.609	3	0.8943	
Task : Domain : Pword syll sec : Bil	0.272	1	0.6017	
Task : Grade : Pword syll sec : Bil	0.300	2	0.8607	
Domain : Grade : Pword syll sec : Bil	2.064	2	0.3563	
Task : Domain : Grade: Pword syll sec : Bil	1.340	2	0.5117	

Control Models for Cognitive Processing (block design test)

Omnibus ANOVA on linear mixed model of STM by Grade with **block design**				
	Chisquare	*Df*	*p*	
Intercept	336.576	1	< 2.2e-16	***
Task	128.764	1	< 2.2e-16	***
Domain	65.816	1	5e-16	***
Grade	41.002	3	6.5e-09	***
blocks	13.849	1	0.0002	***
Task : Domain	0.130	1	0.7187	
Task : Grade	23.652	3	3e-05	***
Domain : Grade	2.288	3	0.5148	
Task : blocks	0.018	1	0.8935	
Domain : blocks	1.498	1	0.2209	
Grade: blocks	13.444	3	0.0038	**
Task : Domain : Grade	4.585	3	0.2048	
Task : Domain : blocks	0.013	1	0.9082	
Task : Grade: blocks	7.038	3	0.0707	.
Domain : Grade: blocks	2.294	3	0.5147	
Task : Domain : Grade: blocks	1.256	3	0.7396	

Omnibus ANOVA on linear mixed model of STM by Group with **block design**				
	Chisquare	*Df*	*p*	
Intercept	369.254	1	< 2.2e-16	***
Task	161.036	1	< 2.2e-16	***
Domain	57.967	1	2.7e-14	***
Group	12.173	2	0.0023	**
blocks	9.899	1	0.0017	**
Task : Domain	0.667	1	0.4142	
Task : Group	19.529	2	5.7e-05	***
Domain : Group	1.050	2	0.5917	
Task : blocks	2.506	1	0.1135	
Domain : blocks	0.132	1	0.7166	
Group: blocks	1.272	2	0.5295	
Task : Domain : Group	3.489	2	0.1748	
Task : Domain : blocks	0.005	1	0.9432	
Task : Group: blocks	3.095	2	0.2128	
Domain : Group: blocks	0.428	2	0.8074	
Task : Domain : Group: blocks	0.367	2	0.8323	

Correlation table between STM measures and reading scores

Correlation table between STM and reading measures overall and for each grade				
	Item verbal	Item nonverbal	Order verbal	Order nonverbal
Across grades				
Word reading errors	-.0323	-.0291	-.0977 ***	-.0328
Word reading speed	.0265	.0493 *	.1170 ***	.0913 ***
Pseudoword reading errors	-.0453 *	-.0023	-.0237	-.0065
Pseudoword reading speed	.0368	.0415 .	.1189 ***	.0902 ***
Grade 1				
Word reading errors	.0454	.0024	-.1881 ***	.0104
Word reading speed	.0814	.0719	.0673	-.0078
Pseudoword reading errors	.0173	-.1346 *	-.0597	.0398
Pseudoword reading speed	.0638	.0926	.0692	.0526
Grade 2				
Word reading errors	-.0301	-.0448	-.0424	.0078
Word reading speed	-.0110	.0066	.0340	-.0313
Pseudoword reading errors	-.0188	.0105	-.0080	.0202
Pseudoword reading speed	-.0152	.0155	.0684 .	-.0015
Grade 3				
Word reading errors	.0346	.0029	-.1004	-.1005
Word reading speed	-.0188	.0225	.2103 ***	.1064 .
Pseudoword reading errors	-.0044	-.0403	-.1165 .	-.0382
Pseudoword reading speed	.1002	-.0049	.1095 .	.0511
Grade 4				
Word reading errors	-.0792 *	-.0188	-.0094	-.0100
Word reading speed	.0529	.0519	.0019	.0794
Pseudoword reading errors	-.0809 *	.0264	.0140	.0144
Pseudoword reading speed	.0655 .	.0196	.0133	.0614

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1