

**This is a preprint of an article whose final and definite form will be published in
Developmental Psychology (in press)**

DOI: [10.1037/a0025660](https://doi.org/10.1037/a0025660) copyright. The American Psychological Association

**This article may not exactly replicate the final version published in the APA journal. It
is not the copy of record.**

Working memory in children: Tracing age differences and special educational needs
to parameters of a formal model

Katrin Göthe, Günther Esser, Anja Gendt & Reinhold Kliegl

University of Potsdam, Germany

Key words: Working memory capacity, interference model, dyslexia, general learning
difficulty

Address for Correspondence:

Katrin Göthe

University of Potsdam

Department of Psychology

Karl-Liebnecht-Str. 24/25

14476 Potsdam

Email: Katrin.Goethe@uni-potsdam.de

Abstract

Parameters of a formal working-memory model were estimated for verbal and spatial memory updating of children. The model proposes interference through feature overwriting and through confusion of whole elements as the primary cause of working-memory capacity-limits. We tested two age groups containing each one group of normal intelligence and one deficit group. For young children the deficit was developmental dyslexia; for older it was a general learning difficulty. The interference model predicts less interference through overwriting but more through confusion of whole elements for the dyslexic children than for their age-matched controls. Older children exhibited less interference through confusion of whole elements and a higher processing rate than young children, but general learning difficulty was associated with slower processing than in the age-matched control group. Furthermore, the difference between verbal and spatial updating mapped onto several meaningful dissociations of model parameters.

Working memory in children: Tracing age differences and special educational needs
to parameters of a formal model

The efficiency of working memory (WM) develops strongly during childhood; it is indispensable for the acquisition of skill and knowledge. Research over the past two decades has established close links between WM and general or special learning difficulties in children. In this study, we attempted to trace differences between normal children and children with special educational needs to differences in theoretically motivated parameters of a formal model of WM (Oberauer & Kliegl, 2006).

WM in children with special and general learning deficits

WM is conceptualized as a system that stores a limited number of independent items and that provides selective access to them for goal-directed processing. Children with special and general learning difficulties show deficits in WM (Gathercole & Pickering, 2001, Exp. 2; Pickering & Gathercole, 2004). Siegel (1994), for example, found that eight-year old children recalled about four items correctly in a complex WM task; eleven-year old children recalled nearly six items. Despite this increase in WM capacity with age, reading-disabled children at the age of eleven performed like normal eight-year old.

Pickering and Gathercole (Pickering & Gathercole, 2004) also reported marked impairments in WM functions for children with special educational needs. Additionally, they identified differential WM profiles for general and specific learning difficulties. Children with specific difficulties in language and literacy exhibited a specific deficit in the verbal domain, whereas children with general learning difficulties were impaired in all dimensions of WM.

Likewise other studies found that, compared to their age-matched controls, children with a general learning difficulty show deficits in many measures of WM (Hasselhorn & Mähler, 2007; Henry & MacLean, 2002; Mähler, 2007; Van der Molen, Van Luit, Jongmans, & Van der Molen, 2007, 2009), whereby the deficits increase with the severity of the learning disability (Henry, 2001; Schuchardt, Gebhardt, & Mähler, 2010). Moreover, special deficits

were also identified for learning disabled compared to their age-matched controls within the verbal domain of WM (Hasselhorn & Mähler, 2007; Henry, 2001; Van der Molen, et al., 2007). Whereas results for the spatial domain were somewhat mixed ranging from worse (Henry & MacLean, 2002; Mähler, 2007) to equal performance, at least in those tasks that only use static and not dynamic aspects of a visual-spatial tasks (Van der Molen, et al., 2009).

In a prospective study, Gathercole and Pickering (2001) screened children at the age of seven and one year later. Those children who were identified as children with special educational needs one year later had performed significantly poorer on WM measures at the age of seven. The findings suggest that poor WM capacity is not only associated with, but a key source for failure in heterogeneous learning activities.

Interference explanations of WM capacity

In the present study, we tested one theoretical proposal about the cognitive architecture of WM and hoped to trace developmental differences as well as learning deficits to different model parameters. Oberauer and Kliegl developed and tested the predictions of their interference model (IM) by applying it to rich data sets collected from young and older adults (Oberauer & Kliegl, 2001, 2006). In the model WM capacity limits arise due to interference and, interference arises due to similarity between elements held active in WM. Increasing the similarity of the to-be-remembered elements decreases memory performance. Aside from model parameters capturing efficiency of processing in the cognitive system, the IM distinguishes two different types of interference: interference through confusion of whole elements and through overwriting of individual features representing each WM element. Confusion of whole elements is at the theoretical core of context models of serial recall (e.g., Burgess & Hitch, 1999) and confusion errors are modulated by factors such as phonological similarity (Henson, Norris, Page, & Baddeley, 1996).

Interference through partial overwriting is adopted from the feature model of Nairne (1990). The model postulates that WM elements are represented by their features. Similar

elements share more features than dissimilar elements. However, one feature cannot be part of the representation of two elements at the same time. A feature that is needed by both representations is assigned to one and gets lost in the other. This overwriting leads to (a) deteriorated representation(s) and to possible forgetting of the respective element(s).

The remainder of the introduction is organized as follows: First we describe the memory-updating tasks used here. Second, we introduce the IM and its parameters. Third, we define age and ability groups of the present sample and fourth specify how we expected that developmental and cognitive-ability differences map onto differences in IM parameters.

Memory-updating tasks

Two memory-updating tasks were administered to the children: a verbal and a spatial one. Figure 1 displays examples of the two tasks. The general design of an updating task involved three phases. First, a variable number of initial starting values was given, whereby the number represented the memory demand (MD), that is, the number of elements to be held and updated in WM. In the spatial task, one or two positions in a three by three grid; in the verbal task, one or two numbers had to be held in WM. Second, the values had to be updated over several successive operations. Updating in the spatial task required shifting of positions; in the verbal task simple arithmetic calculations. Updating operations always applied to the current value in WM with the resulting value serving as input for the next updating. Finally, following the last updating, values were probed for all WM counters.

Complete time-accuracy functions (TAF; Kliegl, Mayr, & Krampe, 1994), covering accuracy between chance and asymptotic performance, were recorded by systematically varying the presentation time (PT) available for an updating operation. Thus, TAFs describe the increase of accuracy with increased PTs. Very short PTs result in close to chance accuracies and longer PTs (i.e., up to several seconds) asymptotically approach maximum accuracy. Hence, we collected data to obtain four TAFs for each subject (2 tasks x 2 MDs). With these TAFs we tested the predictions of the IM as outlined in the next section.

Interference model

In the present article we describe the representational aspects of the IM in a simplified way. We moved model formula to Appendix A. For a detailed description we refer to the original article of Oberauer and Kliegl (2006). The presented cue indicates the element we have to update (e.g., in our spatial task the cue said which animal's position has to be shifted). In order to do this the element has to be activated in the focus of attention (Oberauer, 2002).

A WM element is represented by a set of several active features, that is, in a distributed fashion. This activation pattern binds the feature together (i.e., the features are not active the whole time but fire at a particular beat). Thus, it determines that these features belong to one element. When more than one element is stored in WM, several sets of features are active. Thereby, similar elements share features: the more similar they are the higher this feature overlap.

The IM assumes that at a given time a specific feature can be bound only to one element¹. Thus, feature overlap between any two elements held simultaneously leads to feature overwriting, that is, the loss of a certain proportion of features in each element. This loss degrades their representations. The IM estimates the average feature overlap with a free parameter *C*. When an element in WM is selected for recall or processing the features that represent that element and that are preserved after feature overwriting receive an activation boost. Due to the feature overwriting mechanism, not all features contribute to the activation and hence, maximum activation is not achieved.

A feature that is shared between the target and another element (i.e., a competitor) not only transfers its activation to the target representation in the focus of attention but also to the representation of this competitor. Consequently, also the competitor is minimally activated. Thus, feature overlap not only leads to an activation decrease for the target but also to an increase for a competitor that shares features with the target. In the focus of attention the representation with the highest activation is now selected for further processing. When feature

overlap is not severe, for example, only two elements are stored in WM, the ratio of activations in the focus is generally sufficient to correctly select the actual element for further processing. However, when more elements are held in WM, the probability of feature overlap and, consequently, also of feature overwriting increases. Hence, the number of preserved feature units for each element decreases, which at the end reduces the probability that the correct element receives the highest activation. This dynamic causes the decrease of asymptotic accuracy in TAFs with increasing MD (Oberauer & Kliegl, 2001). The time from presentation of the cue to retrieval into the focus of attention and transformation of that element (i.e., the updating) is estimated with a second free parameter, the processing rate, r .

The second type of interference (i.e., confusion of entire elements) is implemented with the assumption that the filter that assures the exclusive transfer of activation of one specific firing beat (the one for the actual cued element) to the focus of attention is not selective enough. Hence, also activation of other representations is transferred to the focus, which in consequence leads to co-activation of (possible) competitors. This variability in selectivity of the phase filter is estimated as the amount of internal noise of the system, leading to random fluctuations of activations and hence reduced accuracy. The noise parameter, σ , is the third free parameter in the IM.

In its current version, the IM contains actually four free parameters: the average proportion of feature overlap between any two elements in WM, C , the standard deviation of the internal noise, σ , and two processing rates, one for recall from WM in the case of MD-1, r_1 , and another processing rate, r , for MD > 1. The two different processing rates arise due to the fact that for MD > 1 a person has sometimes to switch between two WM elements prior to a processing step. This object switching takes time (Oberauer, 2003). Switching costs do not arise for MD-1. To account for this fact a separate processing rate was introduced for MD-1.

By definition, feature overlap only arises for MD > 1 because only in this case is more than one representation actively held in WM, which is required for sharing features between

them. In our experiment we did not control the amount of feature overlap, but assumed that feature overlap only occurs in the MD-2 condition. Noise, however, is assumed to cause interference in both the MD-1 and MD-2 conditions.

The parameters are free to vary as a function of experimental condition and as a function of individual or group differences. For example, the feature-overlap parameter C should increase with the similarity of the experimental items. In agreement with this prediction, Oberauer and Kliegl (2006, Exp. 2) found that feature overlap for elements coming from both the spatial and verbal domain of WM is smaller than for elements that come from the same domain.

Parameters are also free to vary between groups. For example, it is often assumed that cognitive development in children is associated with an increase in mental speed (e.g., Kail, 1993). This prediction should be reflected in an age difference in the processing rate parameters. In an age-comparative study, Oberauer and Kliegl (2010) showed that older adults exhibited stronger interference through feature overwriting (C parameter) and confusion (noise parameter), but no difference in the processing rate. However, the model estimated a high correlation between noise and processing rate. Hence, age differences could not be specifically assigned to one of the two parameters. To our knowledge, no such model-based test of developmental differences has been carried out for children.

Developmental differences and differences in ability

In the present study we estimated model parameters from WM data of normal children of different ages. Additionally, we examined effects of deviations from normal development on model parameters by contrasting children with and without a learning difficulty. Specifically, we tested four groups of children. Two groups of eight-year old children (second-grade) and two groups of eleven-year old children (fifth- and sixth-grade) were

recruited. This age range had been reported before to cover significant developmental increase in WM functions (for a review see Gathercole, 1998).

At each age level, a group with normal children served as an age-matched control group for another group of children with a cognitive deficit. The control groups included children that were of average intelligence, that is, children's T-scores ranged from 40-59². Children in the young-deficit group suffered from developmental dyslexia (hereafter referred as dyslexia). Children in the old-deficit group were afflicted with a general learning difficulty with an average three-year delay in nonverbal intelligence behind their age norm. This group was intellectually comparable to children of the young control group. Hence, the older children with a general learning difficulty had one control group that was matched in chronological age, and another control group that was matched in mental age.

Linking parameters to developmental effects and to general and specific learning disabilities

Table 1 summarizes our expectations about effects of age, dyslexia and general learning difficulty on the IM parameters. In the case of unspecific hypothesis with respect to one or both interference parameters we included the direction of the expected effect for both.

Developmental effects. With respect to developmental effects, first, we expected older compared to younger children to show faster processing in both domains. Second, we expected that children's WM capacity increases with age in both domains. In the IM, capacity limits may arise due to differences in the feature-overlap or in the noise parameter. We had no specific prediction about whether one or both interference parameters should reflect the expected developmental increase in WM capacity between the two control groups.

Dyslexia effects. The core of the definition of dyslexia lies in a phonological deficit (Frith, 1985; Vellutino, Fletcher, Snowling, & Scanlon, 2004). Therefore, we expected that the dyslexic group shows more interference through feature overlap and confusion than its control group in the verbal domain. There is conflicting evidence as to whether people with

dyslexia are better, same, or worse at spatial processing, with many of the differences being the result of task demands (Jeffries & Everatt, 2004). Several researchers have argued that spatial memory deficits are only apparent when verbal mediation is required by the task (Gould & Glencross, 1990; Thompson, 1982). We assumed that possible effects should be weaker in the spatial than in the verbal domain. Given that spatial updating can hardly be verbally recoded in the present experiment, dyslexia effects may even be absent in this condition. Furthermore, as a tribute to the specificity of the deficit, we expected that dyslexic and control children do not differ in general processing rate.

Age-matched learning difficulty effects. For the children with a general learning difficulty (GLD5/6) we expected the same deficits in both tasks, with lower processing rate and higher susceptibility to interference compared to their age-matched control group (CG5/6). We had no specific expectation whether the deficit would become visible in one or both interference parameters.

Learning difficulty - difference hypothesis. For the comparison of the GLD5/6 group with their intelligence-matched group (CG2) we had two expectations. The first one distinguishes between the difference and the developmental-retardation hypotheses (for a review, see Bennett-Gates & Zigler, 1998). Within the difference hypotheses there are two views: the conventional and the unconventional difference view. The conventional difference hypothesis (Ellis, 1969; Milgram, 1973) postulates that the GLD5/6 group will show worse performance than their age-matched (CG5/6) and their intelligence-matched control groups (CG2). According to this expectation, we should observe group differences in rate and interference parameters (overlap, noise, or both) for both tasks. The unconventional difference hypothesis (Kohlberg, 1968), in contrast, postulates that children with a general learning difficulty should perform better than intelligence-matched children due to their longer learning experience. We did not think that the learning experience would manifest itself in all parameters of the IM, but did not favor a specific prediction.

Learning difficulty - developmental-retardation hypothesis. Seen from the developmental retardation hypothesis (Hodapp, Burack, & Zigler, 1990; Zigler, 1969) learning-disabled children should perform at the same level compared to their intelligence-matched controls. This is because deficits of the GLD5/6 are expected to be quantitative, not qualitative in nature.

Method

Subjects

Children attended schools in Potsdam or neighboring municipalities. Almost all children were drawn from a longitudinal study conducted by the unit of clinical psychology of the University of Potsdam. This study provided baseline data for the present study in order to classify the children as children with normal intelligence or as children with special educational needs. For each child, the BUEGA (Esser, Wyszkon, & Ballaschk, 2008), a screening test for specific developmental disorders in elementary school age (subtests: verbal intelligence, nonverbal intelligence, expressive language, reading, orthography, arithmetic, attention) and further WM tests (word recall forward, digit recall backward, corsi-block task) were administered prior to the memory-updating experiments. In addition, a test of figural short-term memory (Oberauer, 1993) was administered (see Appendix B for a description of all psychometric tests).

We tested four groups of children (total N = 80). We refer to the control group with children of the second grade (n = 20) as CG2, to the dyslexic children of the second grade (n = 21) as DYS2, to the control children of the fifth and sixth grade (n = 20) as CG5/6, and to children with a general learning deficit from the fifth and sixth grade (n = 19) as GLD5/6. The children of the control groups were of average nonverbal intelligence (T-score: 40-59) and did not suffer from any developmental disorder with respect to language problems, dyslexia, dyscalculia or attention deficit hyperactivity disorder. The remaining children of each age

cohort met either the criterion for dyslexia (DYS2) or for general learning difficulty (GLD5/6), both according to the ICD 10 (WHO, 1999).

DYS2 children were in the normal range for nonverbal intelligence, but reading and writing levels were significantly below average of their age group (1.5 SD); they were lower than would be predicted by these children's intelligence scores (1.5 SD). Their impairment in reading and writing development cannot be accounted for with inadequate schooling, social background, or organic causes (e.g., visual acuity problems, or amblyacousia).

GLD5/6 children's nonverbal intelligence ranged from T-scores > 29 (i.e., they were not mentally retarded) and T-scores < 40 (i.e., the lower limit of the age-matched norm). According to BUEGA norms, standardizing the raw scores of the nonverbal intelligence subtest of the GLD5/6 children with the second-grade norms resulted in an average nonverbal intelligence that is ranging between 40 and 59 in T-scores. Thus, their nonverbal intelligence was comparable to that of the two younger groups.

All children had average age-standardized T-scores on attention and arithmetic test, ranging from 40 to 59. This was to ensure that, especially for the two deficit groups (DYS2 and GLD5/6) no additional deficit was diagnosed. Table 2 provides means (SD) of test scores and mean (range) of age for each group.

Data collection lasted almost one year. Second-grade children were tested in summer 2008 and fifth/sixth-grade children in spring 2009. Testing was carried out individually and comprised five to seven one-to-two hour sessions. It took place in a separate room in the school, at home, or in our laboratory. Parental consent was available for all children. Children received 5 EUR for each hour plus extra 5 EUR reward for staying until the last session. This resulted in a standard payment of 30 EUR, but was incremented if more sessions were necessary (see next paragraph). Two dyslexic children dropped out because they were overburdened by the task demands.

General procedure

Three tasks were administered to each child. The tasks included a spatial and a verbal variant of the memory-updating task and the figural short-term memory-task. The figural short-term memory-task was administered either in the first or in the last session (detailed description was moved to Appendix B). In each of the remaining sessions both the verbal and spatial memory-updating experiments were carried out. The order of updating tasks in a session was constant across sessions but counterbalanced across participants. An updating experiment could be terminated prematurely and continued on the consecutive session, if attentional fluctuation was too large. In summary, the order of tasks scheduled for five sessions was adhered to, but could be extended by one or two additional sessions. Average number of sessions was 5.5 for CG2, 5.7 for DYS2, 5.1 for CG5/6, and 5.1 for GLD5/6.

Spatial memory updating. Each updating task (verbal or spatial) comprised 20 blocks of twelve trials each (total N of trials: 240). The spatial task started with presentation of a three by three grid for 500 ms. Subsequently, the image of one animal (either a brown cat or a blue mouse) was displayed in one field of the grid (MD-1). In the MD-2 condition two animals (a brown cat and a blue mouse) were displayed in two different positions within the grid. Presentation order of animals in MD-2 condition was counterbalanced across participants. A press on the space bar triggered the presentation of the position of the other animal together with the disappearance of the first animal's position (see bottom of Figure 1).

With the next key press, the animal disappeared and the updating sequence started with the display of a centrally presented arrow in the color of the to be shifted animal, together with an image of the animal next to the arrow. The arrow could point to one of eight possible directions. The position of the animal had to be mentally shifted one step in the grid based on its actual position along the indicated direction of the arrow, thereby never surpassing the boundaries of the grid (known by the children). In the MD-2 condition the animal position that had to be updated was randomly chosen with the restriction that the

position of each animal was updated at least once in a trial. The arrow presentation and the mental shift formed one updating operation. Three to four updating operations were realized during one trial. The number of updating operations was randomly chosen from trial to trial. The computer-paced PT of the arrow, that is, the time available for one updating operation was chosen from twelve possible predefined PTs (see below for the assignment of these values); the PTs stayed constant during a trial.

After the final updating operation, the position(s) of the animal(s) was/ were probed. Above the empty grid a colored question mark and the image of the respective animal were displayed. The subject clicked with the computer mouse in the field, where the animal was expected. In the MD-2 condition order of probing was the same as order of presenting the initial positions. Whereas intermediate positions could be identical, final positions could not, which was known by the participants.

Verbal memory updating. Verbal updating started with the presentation of an empty basket for 500 ms. In the MD-1 condition, an image of one sort of fruit together with a digit representing the quantity of the fruit were placed into the basket. In the MD-2 condition, two fruits together with the digits representing the respective quantity were displayed sequentially separated by a press on the space bar (see top of Figure 1). Order of presentation was constant across sessions but counterbalanced across participants. The quantity of each fruit ranged between one and nine. After pressing the space bar, the updating sequence started immediately containing three to four updating operations, randomly chosen from trial to trial.

During updating a closed basket was shown together with the updating operation. The updating operation required to “add two” (indicated by a “+2” next to the fruit above the basket) or to “subtract one” (indicated by a “-1” next to the fruit below the basket) from the magnitude of the respective fruit. The children had to compute the quantity of the fruit(s) and had to remember the actual value(s). In the MD-2 condition, each fruit was updated at least once. After the final update, the result(s) of the calculations was / were probed. The basket

disappeared and the image(s) of the fruit(s) was / were shown with a question mark next to it / them. The participant had to type the final value on a numeric keypad. Children knew that all values were always within the range of one to nine. Values were never identical for the final result in the case of MD-2. Feedback consisted of a smiling or a sad face for a correct or a wrong response, respectively.

Presentation times. PTs for a trial were chosen out of twelve possible PTs. PTs ranged between 250 and 1545 ms with an incremental rate of 18% for spatial memory updating, for both MD-1 and MD-2 conditions. For verbal memory-updating conditions, PTs ranged between 515 and 4000 ms with a stepwise increment of 17%. The differences in the selected PT ranges across tasks resulted from specific time courses of the accuracy accrual across tasks (i.e., faster maximum accuracy for the spatial task) which were found suitable based on results of a pilot study with children.

The twelve PTs were divided into three speed classes: fast, medium, and slow with four PTs each. The order of PT-classes was fixed for each block of twelve trials, always repeating a sequence of medium—slow—fast. Within each class, PTs were chosen randomly with the constraint that every PT occurred once in a block. Across each memory-updating experiment (verbal and spatial), this led to ten repetitions of each PT for each MD.

Prior to the first session of each MD of spatial and verbal task, a self-paced variant consisting of six trials was administered in which the child determined the PT for each updating with a press of the space bar. Finally, three practice trials preceded the first test blocks for a given MD; they were identical to the following test trials.

Results

We first describe psychometric data that was available prior to our study and that partly served as selection criteria for our groups. Then, we introduce our strategy of data

analysis and fitting of nonlinear mixed models (NLMMs). Finally, we apply the IM to the data of each child and experimental condition.

Psychometric group profiles

Table 2 summarizes the psychometric data. The data comprise values out of five domains: intelligence (age-matched norms and second grade norms), language, memory, arithmetic and attention. Raw scores are reported for memory tests. For all other tests, we report standardized T-scores (age norms or second grade norms). Analyses of variance (ANOVAs) were computed to describe the four groups and to determine whether selection was appropriate. Table 3 summarizes the results.

In a first step, ANOVAs with age as between-subject factor and dyslexia and learning difficulty as two between-subject contrasts nested within the two age groups were performed for intelligence scores with age-matched norms. Over all, older children scored lower on intelligence tests (total, nonverbal and verbal) than younger children. Obviously, this was due to children with a general learning deficit (GLD5/6) exhibiting lower intelligence scores than their age-matched controls (CG5/6). The two young groups (i.e., CG2 and DYS2) did not differ significantly.

Next, we computed the ANOVA with intelligence scores standardized with second grade norms (which was appropriate for the young children but not for the old). This was done in order to determine whether the GLD5/6 group was comparable in intelligence to the junior groups. The results are summarized in Table 3 (bottom part). Helmert contrasts were specified to compare, first, the mean of the GLD5/6 with the average intelligence of the two young groups. This contrast revealed comparable intelligence scores for the total and the nonverbal intelligence of the young (CG2, DYS2) and the GLD5/6 group. The GLD5/6 outperformed the young groups in verbal intelligence. Hence, the young groups were comparable to each other in intelligence and, as planned, comparable to the GLD5/6 in nonverbal intelligence. The second contrast compared the intelligence scores of the CG5/6

group with the mean intelligence of the other three groups. Applying second-grade norms to intelligence scores, the CG5/6 group, as assumed, outperformed the other groups for all intelligence scores. Taken together the selection of the groups according to intelligence test scores was successful as far as the nonverbal and total intelligence are concerned. The two young groups were comparable in intelligence as were the two young and the GLD5/6 group.

For the other tests we computed ANOVAs with age as a between-subject factor and dyslexia and learning disorder as between-subject factors nested within the two age groups. Let's focus on the language tests first. For expressive language, no age effect was obtained. Interestingly, also dyslexia had no effect on expressive language ability. However, the GLD5/6 group had significantly lower scores than the CG5/6. For reading and orthography tests, young groups scored lower than old groups. This was due to the dyslexic group (DYS2) scoring significantly lower in both tests compared to their control group. These lower reading and orthography scores were completely in agreement with the diagnosis of dyslexia. The GLD5/6 group scored lower than the CG5/6 group in the reading, but not in the orthography test. Hence, the older groups' language scores mirror those for verbal intelligence: The CG5/6 did not only outperform the GLD5/6 group in intelligence, but also in some of the language measures.

On all four memory tests, the older groups outperformed the younger. This result was not surprising as memory scores were not age normalized and the effect reflects the developmental increase in memory capacity. Neither of the deficit factors (dyslexia and learning disorder) was significant, showing that the deficit groups performed equally compared to their control group (CG2 or CG5/6) in all memory tests. This result was unexpected as learning disorders are usually characterized by deficits in memory, but it affords an opportunity to check the diagnostic potential of experimental performance and theoretically motivated model parameters in memory updating.

For scores in attention and arithmetic tests, there was no age, dyslexia or learning difficulty effect. The absence of significant differences in arithmetic performance was welcome, since in our verbal memory task, subjects had to calculate. With these scores, differences in the updating performance cannot be attributed to a generally inferior performance of the deficit groups. Furthermore, we checked that no child had an attention-deficit disorder that could have served as a trivial interpretation of possible group differences in our updating tasks. In summary, we submit that the selection of the groups was successful.

Model testing

Nonlinear increase of accuracy over PTs is typically described with a negatively accelerated exponential function (Kliegl, et al., 1994; McElree & Doshier, 1989); functions were fit separately for each subject and condition, and subsequently function parameters were subjected to standard repeated-measures ANOVAs or t-tests. Applying this method to our experiment we would estimate 320 parameters (four parameters for each of the 80 subjects) for 3840 data points (i.e., twelve data / parameter).

In the present study, following Oberauer and Kliegl (2006), we used nonlinear mixed models (NLMMs) as a framework for data analysis and model fitting (Pinheiro & Bates, 2000). This allows us to describe our results with respect to the effects of experimental conditions on the sample mean (e.g., fixed effects of MD-1 vs. MD-2 or the spatial vs. the verbal task) and with respect to inter-individual differences (i.e., the so-called random effects, that is, the variance associated with these effects across subjects assuming a normal distribution for these effects).

An important advantage of NLMM (e.g., compared to repeated-measures regression analyses) is that it reduces the risk of over-fitting to unreliable differences between individuals. Moreover, we directly estimate the parameters of the IM (along with between-subject variance in the parameters) in a single analysis. In our case, this meant estimating at

most 31 parameters to 3840 data points in the final model (i.e., ~ 124 data / parameter). We used the nlme package (Pinheiro, Bates, DebRoy, & Sarkar, 2005) as provided in the R language and environment for statistical computing (R Development Core Team, 2009).

The data of all subjects were fit simultaneously to the IM. We estimated fixed effects of the experimental and quasi-experimental factors (task: verbal vs. spatial, MD: 1 vs. 2, age: young vs. old, and, the two nested effects within age groups: dyslexia and general learning difficulty) on three model parameters. We carried out several sequences of model building, that is, starting out with different parameters. The model-building sequence reported below is the one that resulted in the best goodness of fit. The general strategy was as follows: We applied a set of successively relaxed versions of the IM. For any significant fixed effect parameter a corresponding random effect (i.e., the between-subject variance of the within-subject effect) was added and maintained if the overall model fit improved significantly according to a log Likelihood Ratio Test. We started with a version (version 0) allowing any fixed effects of the experimental conditions. We then started to test the factors task and MD (not for the C parameter) for each parameter of the IM independently. We integrated the effects and inspected which of them were still significant. Then, the age and the two deficit factors, nested under age, were tested for each parameter. For the final model, we tested whether the estimated correlations between the random effects were significant. If correlations were not significant, they were fixed to zero to simplify model estimations. The history of models and the fit statistics of these model versions are summarized in Appendix C.

Version 20 represented the best fitting model. Figure 2 displays the predictions of the IM together with the observed data as a function of group, MD, and PT. The TAFs reflect the expected effects: (1) Accuracy increases with PT (12), MD (2 vs. 1) leads to TAFs with shallower slopes and lower asymptotes, (3) within the younger children, asymptotic accuracy is a bit lower for the dyslexic than the control group with similar magnitude in the four experimental conditions, (4) within the older children, performance appears to be similar for

children without and with a learning deficit as long as MD is one, but the learning deficit becomes clearly visible when two elements have to be updated. (5) Finally, the drop in asymptotic accuracy is much larger for the spatial than for the verbal task for the learning-deficit children. In summary, the experimental manipulations worked as expected and, by and large, they are also reflected in group differences. In addition to the overall fit indices for version 20 we also calculated R^2_{adj} (see Appendix C for formula) for each group. These were .754, .830, .800 and .765 for the CG2, DYS2, CG5/6 and GLD5/6, respectively. Thus, overall there appears to be no major group difference in this index of goodness of fit; numerically the DYS2 group achieved the highest fit index in version 20.

Group, task, and MD effects on estimates of model parameters

In the following we focus on the group-, task-, and MD-related differences in estimates of IM parameters. Parameter estimates of version-20 are summarized in Table 4; together with the data these model parameters were used to generate the TAFs in Figure 2.

Feature-overlap parameter. In Figure 3 we display conditional modes for the feature-overlap parameter as a function of task and group. In general, conditional modes are predictions of model parameters for each subject conditional on the subject's data and the parameters of the nonlinear mixed model. As these shrinkage-corrected estimates are based on model parameters which were estimated from all data of all subjects, they are not independent observations. The spatial task showed higher feature overlap than the verbal task. The model does not need different feature-overlap parameters for the two age groups, but the model fits best with a separate overlap parameter for the dyslexic group. Somewhat surprisingly, feature overlap is estimated as smaller for the DYS2 group than for the CG2 group in both the verbal and the spatial task. Two additional group effects on feature overlap missed the conventional level of significance. First an age effect appeared to be present that was limited to the spatial task, that is, an interaction of age and task. The effect (although just reaching significance ($t = -1.9$, $p = .05$)) was too small to increase overall model fit (Log-Lik = 3.15, $p = 0.08$). Likewise,

there was a non-significant trend that learning disabled had higher feature overlap compared to their age-matched control group ($t= 1.9$, $p= 0.06$). This trend was observed to the same extent for the verbal and the spatial task and is visible in Figure 3. As it missed the conventional level of significance, we did not keep the parameter in the model.

Noise parameter. The conditional modes for the noise parameter are displayed in Figure 4. Noise was estimated to be larger for the spatial than for the verbal task. For older children lower levels of noise were estimated than for the younger ones. Dyslexia effect on the noise parameter was as expected: Interference through noise was estimated to be larger for dyslexic (DYS2) than for the age-matched control children (CG2) in both tasks. The slightly larger effect for the verbal than for the spatial task ($t = -2.1$, $p = .04$) did not lead to an overall increase in the goodness of fit (Log-Lik = 4.16, $p = 0.124$).

Rate parameter. Figure 5 shows the conditional modes for the rate parameter as a function of group for MD-1 (upper panels) and MD-2 (lower panels) of the verbal (left panels) and the spatial task (right panels). We expected different rates for the two MDs. For MD-2, but not MD-1, we expected a time-consuming object switch. Accordingly, the IM fitted best with different rates for the two MDs. Compare panels in top row and those in bottom row of Figure 5. Furthermore, slower rates were estimated for the verbal than the spatial task (compare values of left and right panels). An increase in MD slowed processing in both tasks, but more so in the spatial tasks. The significant age effect showed that processing was faster for older than for younger children. This effect was more pronounced for the verbal than for the spatial task. Moreover, the GLD5/6 group was slower in processing than their age-matched control group in both tasks.

Parameter correlations. IM fits also include correlations between various random effects. Estimated correlations were at best moderate (see Table 4). The highest correlation, which was negative (-.50) was obtained for the noise parameter in the verbal and the spatial task. Furthermore, feature-overlap parameters for the two tasks were also correlated

negatively (-.35). Finally, the noise parameter in the verbal task correlated weakly negative with processing rate for MD-1 in this task (-.36) and weakly positive with feature overlap in spatial task (+.33). All other correlations were negligibly small.

Finally, we used model parameters and subjects' data to generate predictions for individual participants' TAFs in each of the four conditions. These predictions (along with the data) are documented in Appendices E-H³. Overall, the model parameters recover a remarkable range of task, MD, age, and deficit related effects at the level of individuals.

Relation between psychometric scores and model parameters

How do our experimental results relate to children's psychometric data, especially those that are usually used to define the groups and the special figural short-term memory test? To answer this question, we explored their relation to the conditional modes. We emphasize that this analysis serves only heuristic or exploratory purposes because we only have about 20 subjects in each group and, by definition, conditional modes are not independent observations. We visually inspected these relationships because systematic patterns may guide our future research.

Overall there was no clear-cut picture of a strong dependency between psychometric values and NLMM predictions of individual differences in model parameters. The strongest evidence concerns the relation between figural short-term memory task and model parameters, but only for the older children. High scores in the figural short-term memory task were associated with low feature overlap; noise and fast processing in the verbal task (see Figure 6). These dependencies were not observed for the younger children. We interpret this result as evidence for higher sensitivity of the tests for old than for young children.

Finally, there were several notable relations for the dyslexic children. Low scores in the writing test were associated with higher feature overlap in the verbal task (upper dotplots of Figure 7) whereas lower total intelligence scores were associated with higher feature overlap in the spatial task (middle dotplots of Figure 7); and lower scores in the arithmetic test

were associated with higher verbal noise (lower dotplots of Figure 7). These patterns appear to reflect some task-specific dependencies, but clearly their interpretation would require a follow-up study with larger samples.

Model comparing learning difficulty and intelligence-matched controls

In the overall model described above differences between the GLD5/6 and their intelligence-matched controls (CG2) could not be tested with the particular contrast specification we used. To test differences between GLD5/6 and CG2 we set up a separate test sequence of the IM contrasting only these two groups (Appendix D). The same incremental procedure was applied. Table 5 summarizes the effects of the best fitting version (version 15). This model included the same effects of task and MD as the overall model. Furthermore, age differences were significant only for the rate parameter. The model estimated a higher processing rate for the learning-difficulty group than for the young control group. This result, however, was qualified by an interaction of group with MD: The age advantage in processing rate decreased when two objects had to be updated. This was the case for both tasks. There was a non-significant trend, however, for less noise in the GLD5/6 compared to the CG2 group $t = -1.8$, $p = 0.08$. Thus, even if old children are matched in intelligence with young children, their processing rate is estimated as higher when WM demand is minimal. As soon as an extra load is put onto WM, the processing rate advantage diminishes.

Discussion

Parameters of a formal model of WM, built around two concepts of interference between elements in WM (Oberauer & Kliegl, 2006), were estimated for a verbal and a spatial memory-updating task for four different groups of children. Children differed in age (8 vs. 11 years) and suffered either from dyslexia (half of the younger children) or a general learning difficulty (half of the older children); the other half of the children served as control groups. The IM assumes that limits in WM capacity arise from interference in WM through overwriting of single features (feature-overlap parameter, C) or through confusion of whole

elements (noise parameter, σ). The other two parameters of the IM describe the processing rates when one or two (and more) objects have to be selected and transformed. We summarize results and then discuss their implications for developmental and special-need aspects associated with WM. Tables 1 and 6 summarize our predictions and the results, respectively.

Task and memory-demand effects on model parameters

Our results are in agreement with earlier results that the two tasks lead to differences in feature overlap (Oberauer & Kliegl, 2006). Overall, elements of the verbal task appeared to be better protected against feature overlap than those in the spatial task. Moreover, the noise parameter, promoting interference through confusion of whole elements, was also smaller for the verbal than for the spatial task. As expected, processing rates were lower when two than when one element had to be updated. Processing occurred at a higher rate in the spatial than the verbal task, but the cost associated with an additional element to be tracked was much larger in the spatial than the verbal task. In general, task and memory demand had strong effects in the expected direction on all parameters of the model.

Developmental effects

Psychometric tests. Considering the psychometric data, developmental effects were observed for the memory tests with older children yielding higher scores on all memory tests. This is in line with results of many developmental studies on WM (Gathercole, 1998).

Rate parameter. Processing rate increased with age, which perfectly fits the predictions. This developmental increase was stronger for the verbal than for spatial task. Developmental differences in speed of processing have been regarded in terms of two broad categories of explanation. One view emphasizes global age-related changes (Kail, 1993; Kail & Salthouse, 1994), another emphasizes experiences that lead to changes. Both views are applicable to the overall age effect. However, the differential developmental effect of task material on the rate parameter is rather in accordance with the latter view.

Increased rate of processing with experience in the verbal task could be represented by a shift from performance based on algorithms, which is relatively low, to performance based on rapid, direct retrieval of the appropriate task response (Logan, 1988). In this view, increased rate in the spatial task could also be associated with increased experience, however, less practiced in school than adding and subtracting.

Noise parameter. We were not specific in predicting whether one or both of the two interference parameters cover developmental increase in WM capacity. The results clearly point to a decrease of noise with age. The noise parameter captures interference through confusion of whole elements. There are two possible interpretations of the developmental effect on the noise parameter. First, it can be interpreted that random activation fluctuations are less pronounced for older compared to younger children. Furthermore, noise is assumed to be dependent on the number of possible elements in WM (which was constantly nine in our tasks). Second, the results can be interpreted such that, with age, children can better distinguish between relevant (i.e., possible) and irrelevant information (i.e., elements never appear in the stimulus set) in WM. Thus there are less potential competitors for older children that can be randomly activated. Such a speculation, of course needs further research.

Feature-overlap parameter. We also hoped to observe age effects on feature overlap. Counter to expectation, there was only a trend for a decrease in feature overlap and it was limited to the spatial task; the effect was also too small to increase the overall fit. Developmental differences on the overlap parameter already had been reported by Oberauer and Kliegl (2010) for older ($M = 68.8$) and younger adults ($M = 19.1$) in a numerical-verbal memory-updating task. There were age effects on both feature overlap and noise parameter with more feature overlap and higher levels of noise estimated for older adults. When numerically comparing the estimated means of the feature-overlap parameters for the children (.52) in our study and for the younger (.29) and older adults (.36) in Oberauer and Kliegl (2010), clearly feature overlap is estimated higher for the children compared to the adults.

Such an inference, of course, requires numerous qualifications due to differences in the tasks between our and the Oberauer and Kliegl (2010) experiment (e.g., different format, more updating steps, different PTs). Nevertheless, the direction of estimates for age-related changes in feature overlap from childhood to adolescence is in the expected direction. It is possible that our developmental range was too small to detect changes in feature overlap.

The increase in memory functions with age was documented in many studies for a wide variety of tasks (Gathercole, Pickering, Ambridge, & Wearing, 2004; Isaacs & Vargha-Khadem, 1989; Siegel & Ryan, 1989). Isaacs and Vargha-Khadem (1989) found uniform developmental increases on verbal and spatial span tasks when testing children of seven and 15 years of age with the spatial task constantly lagging about 1.5 items behind. This is in line with our results on the noise parameter for which we found similar age effects for the verbal and spatial task. Generally, performance on short term and WM tasks was found to steeply increase up to the age of ten; thereafter a more gradual improvement was discovered with an asymptotic level reached at the ages of eleven or twelve years (Gathercole, 1999). An exception represents performance on the listening span task, a WM task for which a steep slope is visible up to the age of 16 (Siegel, 1994). This could point to a different developmental profile of WM in contrast to short term memory tasks for which a longer period of development can be possibly assumed. The latter result would be in line with neuropsychological findings since also frontal lobe functions strongly associated with WM (Smith & Jonides, 1998) show such a enduring developmental time course (Diamond, 1990).

Within the IM developmental effects on WM functions, that is, storage (feature overlap, noise) and processing functions (rate) were modeled independently. There are other accounts postulating a general limited resource that must be shared between storage and processing, that is, assuming a dependency of both (Case, Kurland, & Goldberg, 1982; Just & Carpenter, 1992; Pascual-Leone, 1970). Pascual-Leone, for example, assumes that developmental change proceeds as a change in sharing ratio between storage and processing,

that is, as processing gets faster resource is freed and used for storage functions. In a precursor publication to the IM Oberauer and Kliegl (2001) tested several competing accounts. Resource accounts failed to explain the data. The problem for such models was that they predicted a pronounced drop in accuracy from MD-1 to MD-2 followed by a smaller drop for higher MDs. Data, however, showed an accelerated drop in accuracies with increasing MD.

Dyslexia effects

Psychometric tests. Per definition, dyslexic children were selected according to their intelligence (average), reading and writing (below average) scores. This resulted in a typical sample of dyslexic children with two exceptions. First, dyslexic children were comparable to their controls with respect to the scores in the expressive language test, which was a grammar test. Furthermore, psychometric memory tests revealed no differences for the dyslexic relative to their age-matched control group. This was surprising to us because a lot of studies revealed deficits in WM for dyslexic children. In contrast to the psychometric tests, our memory updating data and subsequent IM parameter estimations unveiled very clear group-differential WM-related effects. Possibly, the experimental assessment is more sensitive than the psychometric assessment in this case.

Rate parameter. There were no dyslexia effects on the rate parameters. As we predicted, there was no need to assume an effect of dyslexia on processing rate when linking nonverbal intelligence (which did not differ between control and dyslexic group) to mental speed (Kail, 1993; Kail & Salthouse, 1994).

Noise parameter. We expected effects of dyslexia on the interference parameters. Indeed, noise level was estimated higher for the dyslexic in comparison to their control group. Counter to expectation, this effect was similar for verbal and the spatial tasks, although we had expected the effect to be absent or at least smaller in the spatial task. There was one version of the IM test sequence in which the noise parameter for the dyslexic group was

estimated as lower in the spatial than in the verbal task, but this effect did not lead to an overall increase in goodness of fit and therefore was not included in the model. Nevertheless, the qualitative pattern also held for the spatial task.

In general, research on dyslexia has emphasized specific deficits in the verbal domain, but recent research also highlighted some effects in the spatial domain. Associated results range from disadvantages for dyslexic (Helland & Asbjørnsen, 2003; Olson & Datta, 2002; Winner et al., 2001) over non-significant differences (Gould & Glencross, 1990; Jeffries & Everatt, 2004; Kibby, Marks, Morgan, Long, & Long, 2004) to advantages for the dyslexic groups (Von Károlyi, Winner, Gray, & Sherman, 2003). The disadvantages in spatial WM for dyslexic compared to normal children are typically related to general problems in executive control for dyslexic (Winner, et al., 2001) or a dysfunctional verbal recoding strategy of nonverbal content (Gould & Glencross, 1990), not due to a global problem of visual-spatial processing. According to this perspective, a possible explanation for our results is that the differences are due to the executive load associated with our WM task.

There are other theories that relate dyslexia to a visual perceptual deficit. In a recent article Vidyasagar and Pammer (2009), for example, argued that the causal deficit in dyslexia is poor visual coding. The authors claim that deficits in attentional mechanisms supervising serial scanning of letters cause a cascade of effects leading to impairment in visual processing of graphemes, their translation into phonemes and subsequent development of phonological awareness. In this view the observed deficit in the spatial task for the dyslexic compared to their controls would be explained by this general perceptual deficit. Evidence for differences should be sufficient motivation for further investigation of these effects. They are of much relevance for a theoretical understanding of dyslexia.

Feature-overlap parameter. We had predicted that dyslexic children will exhibit more interference on noise and feature-overlap parameter (Table 1). However, within the confines of the IM, feature overlap and hence overwriting appears to happen less strongly for

dyslexic than for control children. This result is surprising. The reason can be found in the characteristics of the observed TAFs for the dyslexics and the mathematical implementation of the feature-overlap parameter. Dyslexic children (who scored lower than their age-matched controls in the condition with low memory demand) showed a smaller drop in accuracies due to memory demand, that is, from MD-1 to MD-2 compared to their age-matched control group. In the present formulation of the IM the feature-overlap parameter captures the drop from MD-1 (where no overlap is present) to MD-2 (where overlap between two elements is present). Hence, the unexpected result in the TAFs transferred to model parameters.

There are several possible explanations for this surprising result in the TAFs. One of them is an effect of sampling. As described in the result section, dyslexic children had lower scores in reading and writing tests than their age mates, but they did not differ in expressive language. This latter result was unexpected and, hence, may be a hint that we possibly selected a special group of dyslexic children.

An alternative explanation, supported by inspection of the data (Figure 2), is that a floor effect masked the effect of storage demand in the dyslexic group. One argument against this is that we scored responses as correct only when both elements were updated correctly. The conditional probability to correctly recall two digits is lower than .7. Hence, even accuracies in high memory-demand conditions were above chance for both the dyslexic and the age-matched controls. Moreover, according to the floor-effect interpretation, the same level of accuracy should hold for shorter and longer PTs.

Such an unexpected result has several implications. One would be the attempt of a replication and extension (e.g., with MD-3). This is important since results on the feature-overlap parameter for the learning-difficulty group and for the age effect, although not significant, pointed into the predicted direction. Another consequence is a reconsideration of the implementation of the feature-overlap parameter. As we stated in the introduction, the application of the IM to the present data was a further evaluation of the model and its

parameters. Therefore, these results may guide us towards a more appropriate specification of the IM.

Learning difficulty effects

Psychometric tests. Children with a general learning difficulty were selected according to nonverbal intelligence scores (below average). Hence, on this score the learning-difficulty group performed worse than the age-matched but equal compared to the intelligence-matched control group. Significantly lower scores compared to the age-matched controls were, moreover, observed in verbal intelligence, reading, and expressive language tests. This further validates the group selection. In contrast, no learning difficulty effect could be observed for the memory tests. This points to a non-typicality of this group given the research results on learning difficulties and associated WM deficits presented in the introduction (Hasselhorn & Mähler, 2007; Henry, 2001; Henry & MacLean, 2002; Van der Molen, et al., 2007, 2009). Moreover, concerning the verbal intelligence score, learning-disabled children outperformed their intelligence-matched controls. Over all (total intelligence) this advantage, however, did not reach significance. We discuss the potential effects of group selection on parameter estimations below.

Rate parameter. Learning difficulty significantly decreased processing rate compared to age-matched controls. The rate deficit was similar for the verbal and the spatial task. This result mirrors the result for nonverbal intelligence and is interpreted as lower mental speed for learning-disabled compared to age-matched children. Rates were significantly higher for the learning-difficulty group when compared to their intelligence-matched controls. This advantage may be partly due to somewhat higher verbal intelligence scores of the learning-disability compared to the intelligence-matched group although groups did not differ in nonverbal and total intelligence. There is another qualification related to WM demand. The processing-rate advantage for the learning-disabled over the intelligence-matched group was smaller when two elements had to be updated. The difference between rate for MD-1 and

MD>1 is linked to the switch between two elements in WM. Such object switch costs are the topic of research on selective attention. The advantage of the learning-difficulty group in rate was probably limited to the activation rate in the absence of executive-control processes.

Noise parameter. There was no evidence for a difference in the noise parameters of the two older groups of children. When the learning-difficulty group was compared to the intelligence-matched control group, we obtained numerical but non-significant evidence for less noise for the learning-difficulty group. The latter result may point to a further developmental advantage of the learning-difficulty group compared to their intelligence-matched controls, which, however, did not bear statistical testing.

Feature-overlap parameter. There was a non-significant trend that learning-disabled children had higher feature overlap than their age-matched controls. No such trend was observable when comparing learning disabled with their intelligence-matched controls.

Conventional difference hypothesis. Compared to the younger children the learning disabled did not exhibit any disadvantages. Hence, there was no evidence for the conventional difference hypothesis, according to which the learning-disabled group should not only score lower compared to their age-matched but also to their intelligence-matched controls.

Unconventional difference hypothesis. However, concerning the model parameters there was evidence for the unconventional difference position that the learning-difficulty group performed better than the young control group with respect to the processing rate. Henry and MacLean (2002) reported higher scores for a learning-difficulty group compared to intelligence-matched controls for a listening span task. They explained this with a change in strategy use and/ or higher familiarity of the verbal material for the learning difficulty compared to their intelligence-matched controls. Similarly, in our experiment older learning deficit children may have had more practice, especially with calculating (in the verbal task) compared to the younger children. This allows them a direct retrieval of results instead of online processing. (We used this explanation also to interpret the age effect on rate.)

Developmental retardation hypothesis. There was much support for developmental retardation: Children in the learning-difficulty group performed at the same level as the young control group with respect to feature overlap and the noise in both tasks. Since there was an age effect on the noise parameter, the absence of it for the CG2: GLD5/6 comparison supports the developmental view. Taken together, we did not obtain support for a structural deficit but rather for a developmental-delay account with a special strength in rate of processing for the GLD group (e.g., when, overlearned simple arithmetic was needed).

Sample selection

The lack of observed group differences in the psychometric memory tests might indicate a non-typical selection of children with learning difficulties since memory deficits for these groups have been reported before (Gathercole & Pickering, 2000, 2001; Hasselhorn & Mähler, 2007; Henry & MacLean, 2002; Van der Molen, et al., 2007, 2009). Moreover, the lack of a difference between the dyslexic and their control group with respect to the expressive language is an evidence for the non-typicality. Although we had access to a very large sample to select from ($N = 2300$) and sample selection was done with extraordinary care and proficiency, the recruitment process is not immune to sampling errors. This applies also to our study since we could only pseudo-randomly select the children, particularly those of the learning-difficulty groups. These samples were pretty much restricted. The total number of children with an a priori diagnosed general learning difficulty, for example, hardly exceeded the number of children that consented to participate. This could have influenced our results so that group differences were less pronounced than predicted. Fully in line with the predictions, the non-significant trends in parameter estimates (i.e., higher feature overlap for GLD5/6 compared to CG5/6, lower noise in the spatial than in the verbal task for the DSY2 compared to CG2), therefore, may be interpreted as a hint for further group differences.

Convergence of psychometric data, experimental data and mathematical modeling

In the ideal world, developmental and deficit groups established by age and psychometric criteria based on scores in intelligence tests and tests of specific abilities map unambiguously on differences in parameter estimates. The IM does not live up to this expectation completely, but there were some interesting leads. For example, for the dyslexic the lower writing scores were associated with higher feature-overlap parameters. However, our results about the relation between psychometric profiles and model parameters were not as clear as expected. As usual there are many reasons for such a relative lack of agreement. Here are two obvious ones. First, NLMM-based parameters represent a step forward in a theory-guided foundation of developmental and disability differences and due to the shrinkage correction inherent to this approach, they are superior to traditional within-subject estimates. Nevertheless, they still may fall short of the reliability of traditional psychometric measures. Second, although the IM recovered the TAF profiles remarkably well, the parameter estimates may point to serious limits of the IM. There are other theoretical ideas (e.g., inhibitory deficit theories) that were not put to test with this model. Thus, the framework of research and modeling implemented could serve as a blueprint for future research.

Nonlinear mixed model framework

In the present article, we offer a view on dyslexia and learning disability from an experimental perspective using a formal model of WM, and estimating parameters with NLMM. Traditionally, these issues are addressed from a psychometric perspective. This perspective puts much weight on defining latent constructs (e.g., in the context of structural equation models) to reduce the error variance associated with fallible indicators. The NLMM perspective reduces error in fallible indicators of subject-based “estimates” of model parameters. Error reduction is based on shrinkage of within-subject based estimates relative to fixed effects and variance components estimated from the complete data set simultaneously (i.e., the conditional modes). At this point in time, there are a few attempts to merge the two

approaches in the context of linear mixed models, but it will probably still be some time before analogous statistics become available for parameters estimated in NLMM.

Summary

Some of the highlights of the present experiment were the following. First, we were able to determine complete time-accuracy functions for children of different age and deficit groups. This by itself is no small achievement. Moreover, we applied a theoretical WM model to these functions and related its formal model parameters to differences in age, reading, and learning ability. With age faster processing and lower interference due to noise was found. Dyslexia was associated with higher noise. One unexpected result was that dyslexia showed an advantage with respect to interference through feature overwriting in both domains. Future work has to show whether such a result can be replicated and extended (e.g., with higher memory demands). This result, moreover, may point to a weakness in the formulation of the feature-overlap parameter within the model framework. The general learning-difficulty group showed faster processing than an intelligence-matched but slower processing than an age-matched control group. There were also several non-significant group trends in line with the theoretical assumptions. This is encouraging because the experiment demonstrated the feasibility of this approach to merge complex experimental research, mathematical modeling, and research on developmental and disability differences.

References

- Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah: NJ: Erlbaum.
- Bennett-Gates, D., & Zigler, E. (1998). Resolving the developmental-difference debate: An evaluation of the triarchic and systems theory models. In A. J. Burack, R. M. Hodapp & E. Zigler (Eds.), *Handbook of mental retardation and development* (pp. 115-131). Cambridge: Cambridge University Press.
- Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*, 551-581.
- Case, R., Kurland, M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*(386-404).
- Diamond, A. (1990). Developmental time course in human infants and infant monkeys, and the neural bases of, inhibitory control in reaching. *Annals of The New York Academy of Sciences*, *608*, 394-426.
- Ellis, N. R. (1969). A behavioral research strategy in mental retardation: Defense and critique. *American Journal of Mental Deficiency*, *73*, 557-567.
- Esser, G., Wyszkon, A., & Ballaschk, K. (2008). *BUEGA: Basisdiagnostik Umschriebener Entwicklungsstörungen im Grundschulalter [Basic diagnostics of specific developmental disorders in elementary school age]*. Göttingen: Hogrefe.
- Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. E. Patterson, J. C. Marshall & M. Coltheart (Eds.), *Surface Dyslexia* (pp. 301-330). London: Routledge & Kegan Paul.
- Gathercole, S. E. (1998). The development of memory. *Journal of Child Psychology and Psychiatry*, *39*(1), 3-27.
- Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, *3*(11), 410-419.

- Gathercole, S. E., & Pickering, S. J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Educational Psychology, 70*, 177-194.
- Gathercole, S. E., & Pickering, S. J. (2001). Working memory deficits in children with special educational needs. *British Journal of Special Education, 28*, 89-97.
- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*, 177-190.
- Gould, J. H., & Glencross, D. J. (1990). Do children with a specific reading disability have a general serial-ordering deficit? *Neuropsychologica, 28*(3), 271-278.
- Hasselhorn, M., & Mähler, C. (2007). Phonological working memory of children in two german special schools. *International Journal of Disability, Development and Education, 54*(2), 225-244.
- Helland, T., & Asbjørnsen, A. (2003). Visual-sequential and visuo-spatial skills in dyslexia: Variations according to language comprehension and mathematics skills *Child Neuropsychology, 9*(3), 208-220.
- Henry, L. (2001). How does the severity of a learning disability affect working memory performance? *Memory, 9* 233-247.
- Henry, L., & MacLean, M. (2002). Working memory performance in children with and without intellectual disabilities. *American Journal on Mental Retardation, 107*, 421-432.
- Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology, 49A*, 80-115.
- Hodapp, R. M., Burack, J. A., & Zigler, E. (1990). The developmental perspective in the field of mental retardation. In R. M. Hodapp, J. A. Burack & E. Zigler (Eds.), *Issues in the*

- Developmental Approach to Mental Retardation* (pp. 3-26). Cambridge University Press.
- Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*(3), 427-466.
- Isaacs, E. B., & Vargha-Khadem, F. (1989). Differential course of development of spatial and verbal memory span: A normative study. *British Journal of Developmental Psychology*, *7*, 377-380.
- Jeffries, S., & Everatt, J. (2004). Working memory: Its role in dyslexia and other specific learning difficulties. *Dyslexia*, *10*, 196-214.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122-149.
- Kail, R. (1993). Processing time decreases globally at an exponential rate during childhood and adolescence. *Journal of Experimental Child Psychology*, *56*, 254-265.
- Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, *86*, 199-225.
- Kibby, M. Y., Marks, W., Morgan, S., Long, S., & Long, C. J. (2004). Specific impairment in developmental reading disabilities. *Journal of Learning Disabilities*, *37*(4), 349-363.
- Kliegl, R., Mayr, U., & Krampe, R. T. (1994). Time-accuracy functions for determining process and person differences: An application to cognitive aging. *Cognitive Psychology*, *26*, 134-164.
- Kohlberg, L. (1968). Early education: A cognitive-developmental view. *Child Development*, *39*, 1013-1062.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492-527.
- Mähler, C. (2007). Arbeitsgedächtnisfunktionen bei lernbehinderten Kindern und Jugendlichen [Working memory functions in learning-disabled children and

- adolescent]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 39(2), 97-106.
- McElree, B., & Doshier, B. A. (1989). Serial position and set size in short-term memory: the time course of recognition. *Journal of Experimental Psychology: General*, 118, 346-373.
- Milgram, N. A. (1973). Cognition and language in mental retardation: Distinctions and implications. In D. K. Routh (Ed.), *The experimental psychology of mental retardation* (pp. 157-230). London: Crosby Lockwood Staples.
- Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, 18, 251-269.
- Oberauer, K. (1993). Die Koordination kognitiver Operationen. Eine Studie zum Zusammenhang von 'working memory' und Intelligenz [The coordination of cognitive operations. A study on the relationship of working memory and intelligence]. *Zeitschrift für Psychologie*, 201, 57-84.
- Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28, 411-421.
- Oberauer, K. (2003). Selective attention to elements in working memory. *Experimental Psychology*, 50, 257-269.
- Oberauer, K., & Kliegl, R. (2001). Beyond resources: formal models of complexity effects and age differences in working memory. *European Journal of Cognitive Psychology*, 13, 187-215.
- Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, 55(4), 601-626.

- Oberauer, K., & Kliegl, R. (2010). Interferenz im Arbeitsgedächtnis: Ein formales Modell [Interference in working memory: A formal model]. *Psychologische Rundschau*, *61*(1), 33-42.
- Olson, R., & Datta, H. (2002). Visual-temporal processing in reading-disabled and normal twins. *Reading and Writing: An Interdisciplinary Journal*, *15*, 127-149.
- Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica*, *63*, 301-345.
- Pickering, S. J., & Gathercole, S. E. (2004). Distinctive working memory profiles in children with special educational needs. *Educational Psychology*, *24*(3), 393-409.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer Verlag.
- Pinheiro, J. C., Bates, D. M., DebRoy, S., & Sarkar, D. (2005). nlme: linear and nonlinear mixed effect models (Version 3.1-96).
- R Development Core Team. (2009). R: A Language and Environment for Statistical Computing (Version 2.9). Vienna, Austria: R Foundation for Statistical Computing.
- Raffone, A., & Wolters, G. (2001). A cortical mechanism for binding in visual working memory. *Journal of Cognitive Neuroscience*, *13*, 766-785.
- Schuchardt, K., Gebhardt, M., & Mähler, C. (2010). Working memory functions in children with different degrees of intellectual disability. *Journal of Intellectual Disability Research*, *54*(4), 346-353.
- Siegel, L. S. (1994). Working memory and reading: A life-span perspective. *International Journal of Behavioral Development*, *17*(1), 109-124.
- Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development*, *60*, 973-980.

- Smith, E. E., & Jonides, J. (1998). Neuroimaging analyses of human working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 12061-12068.
- Thompson, M. E. (1982). The assessment of children with specific reading difficulties (dyslexia) using the British Ability Scales. *The British Journal of Psychology*, 73(4), 461-478.
- Van der Molen, M. J., Van Luit, J. E. H., Jongmans, M. J., & Van der Molen, M. W. (2007). Verbal working memory in children with mild intellectual disabilities. *Journal of Intellectual Disability Research*, 51(2), 162-169.
- Van der Molen, M. J., Van Luit, J. E. H., Jongmans, M. J., & Van der Molen, M. W. (2009). Memory profiles in children with mild intellectual disabilities: Strengths and weaknesses. *Research in Developmental Disabilities*, 30, 1237-1247.
- Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, 45(1), 2-40.
- Vidyasagar, T. R., & Pammer, K. (2009). Dyslexia: A deficit in visuo-spatial attention, not in phonological processing. *Trends in Cognitive Sciences*, 14(2), 57-63.
- Von Károlyi, C., Winner, E., Gray, W., & Sherman, G. F. (2003). Dyslexia linked to talent: Global visual spatial ability. *Brain and Language*, 85, 427-431.
- WHO. (1999). *International Classification of Diseases. ICD-10 Chapter V*. Bern: Hans Huber.
- Winner, E., von Károlyi, C., Malinsky, D., French, L., Seliger, C., Ross, E., et al. (2001). Dyslexia and visual-spatial talents: Compensation vs deficit model. *Brain and Language*, 76, 81-110.
- Zigler, E. (1969). Developmental versus difference theories of mental retardation and the problem of motivation. *American Journal of Mental Deficiency*, 73, 536-556.

Author Note

This research was funded through Deutsche Forschungsgemeinschaft (Grant KL 955/16-1). Data and R scripts for analyses and figures reported in this article are available at the Potsdam Mind Research Repository (PMR2; <http://read.psych.uni-potsdam.de/pmr2/>). We invite replication and alternative model specifications and tests.

We thank Petra Grüttner, Ireen Saal, Elizabeth Jünger, Isabel Lima Pimentel, Paul Endrejat, Nadine Poltz, Janine Schäferhoff and Juliana Seidler for data collection and subject supervision. The authors thank the school, the parents, and the children who consented to participate in this study. We also thank Klaus Oberauer for thoughtful comments on earlier versions of the paper. In addition, we thank Florian Schmiedeck and two anonymous reviewers for their useful comments.

Footnotes

- 1) The logic is adapted from synchronized firing units (Hummel & Holyoak, 1997; Raffone & Wolters, 2001) and further elaborated in Oberauer and Kliegl (2006).
- 2) We report ranks on the T- scale instead of the IQ- scale. The T- scale has a mean of 50 and a Standard Deviation (SD) of 10. T values can be converted into IQ values. A T value of 50 represents an IQ of 100.
- 3) At the individual level some fits look less than optimal for some participants. There are two reasons. First, of course model predictions of the best fitting version are not perfect due to limits of the IM and due to the fact that a non-saturated version was implemented (trading variance explanation in parsimony). Second, the NLMM corrects for unreliable individual differences by shrinkage correction. This implies that means that (a) are more extreme, (b) are represented by a smaller number of observations, or (c) have a larger within-subject variance “borrow” strength from the more reliable population mean, that is they are shrunken towards the population mean.

Table 1: Predictions of the IM for the group differences in parameter profiles.

Effect	Task	VMU			SMU		
	Contrast/ Parameter	overlap	noise	rate	overlap	noise	rate
Age	CG5/6 vs. CG2	+	+	+	+	+	+
DYS	CG2 vs. DYS2	+	+	0	0	0	0
GLD	CG5/6 vs. GLD5/6	+	+	+	+	+	+
	CG2 vs. GLD5/6 (CDH)	+	+	+	+	+	+
	CG2 vs. GLD5/6 (UDH)	-	-	-	-	-	-
	CG2 vs. GLD5/6 (DRH)	0	0	0	0	0	0

Note. A + represents more efficient performance for the respective first group, a - represents less efficient performance for the respective first group, 0 represents no difference. **DYS:** dyslexia, **GLD:** general learning difficulty, **VMU:** verbal memory-updating task, **SMU:** spatial memory-updating task, **CG2:** control group 2nd grade, **DYS2:** dyslexic group 2nd grade, **CG5/6:** control group 5th/ 6th grade, **GLD5/6:** general learning-difficulty group 5th/ 6th grade, **CDH:** conventional difference hypothesis, **UDH:** unconventional difference hypothesis, **DRH:** developmental retardation hypothesis.

Table 2. Psychometric data for the four sample groups.

Domain	DV	CG2 n = 20	Dys2 n = 21	CG5/6 n = 20	GLD5/6 n = 19
	Age (years, month)	7.7 (0.5)	8.0 (0.6)	11.10 (0.8)	11.10 (0.7)
Int (age matched norm)	total	50.25 (5.3)	52.57 (7.16)	52.85 (3.88)	37.21 (2.53)
	nonverbal	52.80 (7.08)	53.86 (8.34)	54.75 (4.85)	38.42 (6.54)
	verbal	48.70 (8.03)	51.48 (6.68)	50.75 (6.39)	40.95 (7.52)
Language	Expressive	50.05 (4.26)	51.19 (6.97)	51.00 (5.37)	46.11 (8.61)
	Reading	50.05 (5.42)	35.76 (5.29)	50.40 (5.78)	46.16 (8.29)
	Orthography	48.20 (5.10)	32.33 (9.51)	48.90 (4.18)	47.26 (6.98)
Memory	Figural STM	68.44 (5.18)	68.08 (6.70)	74.95 (5.63)	72.23 (6.50)
	Word recall forward	4.05 (0.69)	3.95 (1.02)	4.65 (0.75)	4.47 (0.70)
	Digit recall backward	3.30 (0.73)	3.10 (0.83)	4.00 (1.08)	3.58 (0.83)
	Corsi-block (simple)	5.30 (1.26)	5.62 (1.20)	6.60 (1.23)	6.37 (1.16)
	Corsi-block (complex)	4.35 (1.14)	4.00 (1.58)	5.30 (0.73)	5.32 (0.75)
	Arithmetic	51.50 (6.66)	48.86 (7.74)	48.95 (4.88)	47.53 (5.35)
	Attention	50.75 (5.25)	50.52 (5.19)	51.90 (3.80)	50.00 (6.03)
Int (2 nd grade norm)	total			66.95 (4.80)	53.74 (3.78)
	nonverbal			63.35 (4.92)	50.11 (6.97)
	verbal			67.10 (6.82)	57.32 (7.65)

Note. First column: domain of test, Second column: dependent variable (DV). Mean score for each group (CG2: control group 2nd grade, DYS2: dyslexic group 2nd grade, CG5/6: control group 5th/ 6th grade, GLD5/6: general learning-difficulty group 5th/ 6th grade) is given together with the Standard Deviation in brackets. The intelligence values (Int) for the 2nd grade and age-matched norm equal each other for CG2 and DYS2 groups, Figural STM: figural short-term memory task, Description of the psychometric tests is given in the Appendix.

Table 3. Results of the ANOVAs results on the scores of the psychometric data.

Domain	DV	Age contrast	DYS contrast	GLD contrast
Int (age matched norm)	total	29.86, < 0.001	2.14, 0.148	92.38, < 0.001
	nonverbal	6.676, 0.012	1.53, 0.219	18.18, < 0.001
	verbal	18.28, < 0.001	<1, 0.622	55.423, < 0.001
Language	Expressive	1.94, 0.168	<1, 0.575	5.55, 0.021
	Reading	15.94, < 0.001	53.16, < 0.001	4.46, 0.038
	Orthography	27.91, < 0.001	55.85, < 0.001	<1, 0.454
Memory	Figural STM	16.00, < 0.001	<1, 0.938	1.99, 0.163
	Word recall forward	9.83, 0.002	<1, 0.699	<1, 0.496
	Digit recall backward	9.32, 0.003	<1, 0.458	2.24, 0.139
	Corsi-Block (simple)	14.17, < 0.001	<1, 0.404	<1, 0.554
	Corsi-Block (complex)	20.70, < 0.001	1.01, 0.319	<1, 0.965
	Arithmetics	1.80, 0.183	1.81, 0.183	<1, 0.482
	Attention	<1, 0.767	<1, 0.888	1.34, 0.250
			GLD5/6 vs. CG2, DYS2	CG5/6 vs. GLD5/6, CG2, DYS2
Int (2 nd grade norm)	total		2.98, 0.088	110.22, < 0.001
	nonverbal		2.43, 0.123	38.12, < 0.001
	verbal		13.70, < 0.001	59.89, < 0.001

Note. The first column includes the broad domain of the dependent variable (DV), which is listed in the second column. The last three columns contain the F-and p-values of the ANOVA computed for the respective contrast. For significant results the p-values are written in bold font. Contrasts for the intelligence scores with the 2nd grade norms are different to the

contrasts computed for all other DV, DYS: dyslexia, GLD: general learning difficulty, CG2: control group 2nd grade, DYS2: dyslexic group 2nd grade, CG5/6: control group 5th/ 6th grade, GLD5/6: general learning-difficulty group 5th/ 6th grade. Description of the psychometric tests is given in the Appendix.

Table 4: Parameter estimates of the Interference model (version 20)

<i>Effect</i>	<i>Mean</i>	<i>SD</i>	<i>T (p-value)</i>	<i>Intercept of C</i>	<i>Task of C</i>	<i>Intercept of r</i>	<i>Intercept of σ</i>
<i>Intercept of C</i>	0.52	0.16	11.23 (<.001)				
<i>Task of C</i>	0.36	0.27	8.11 (<.001)	-0.35			
<i>Age of C</i>	-0.07	fixed	-1.42 (0.156)				
<i>Dys of C</i>	-0.12	fixed	-2.20 (0.028)				
<i>Intercept of r</i>	1.48	0.33	19.53 (<.001)	-0.10	0.14		
<i>Task of r</i>	2.93	fixed	17.38 (<.001)				
<i>MD of r</i>	-0.55	fixed	-8.15 (<.001)				
<i>Age of r</i>	0.89	fixed	6.09 (<.001)				
<i>GLD of r</i>	-0.39	fixed	-2.33 (0.020)				
<i>Task* MD of r</i>	-1.51	fixed	-6.96 (<.001)				
<i>Task* Age of r</i>	-0.62	fixed	-3.08 (0.002)				
<i>Intercept of σ</i>	0.22	0.03	30.73 (<.001)	0.13	0.33	-0.36	
<i>Task of σ</i>	0.04	0.03	8.01 (<.001)	0.15	-0.04	0.08	-0.50
<i>Age of σ</i>	-0.02	fixed	-2.05 (0.040)				

<i>Dys of σ</i>	0.03	fixed	3.17 (0.002)				
-----------------------------------	------	-------	--------------	--	--	--	--

Note. *C*: feature-overlap parameter, *r*: rate parameter, σ : noise parameter, task: task effect, MD: effect of memory demand, Age: effect of age, Dys: effect of dyslexia (nested under younger group), GLD: effect of general learning difficulty (nested under older age group). The table contains estimates for the parameter means (Mean) and their standard deviations (SD) together with the T values (p-values are given in brackets). In the right part of the table (column six to nine) estimates for the parameter correlations across participants are given. Estimates are based on 3840 data points.

Table 5: Parameter estimates of the Interference model for the GLD5/6- CG2 comparison (version 15)

<i>Fixed effect</i>	<i>Mean</i>	<i>SD</i>	<i>T (p-value)</i>	<i>Intercept of C</i>	<i>Intercept of r</i>	<i>Intercept of σ</i>
<i>Intercept of C</i>	0.38	0.21	7.84 (<.001)			
<i>Task of C</i>	0.48	fixed	11.07 (<.001)			
<i>Intercept of r</i>	1.52	0.40	12.88 (<.001)	0.12		
<i>Task of r</i>	2.59	fixed	12.69 (<.001)			
<i>MD of r</i>	-0.68	fixed	-9.11 (<.001)			
<i>DiffDev of r</i>	0.64	fixed	3.34 (<.001)			
<i>Task* MD of r</i>	-1.18	fixed	-4.05 (<.001)			
<i>Task* DiffDev of r</i>	-0.32	fixed	-2.48 (0.013)			
<i>Intercept of σ</i>	0.21	0.03	34.64 (<.001)	0.30	-0.15	
<i>Task of σ</i>	0.04	0.14	5.58 (<.001)	0.36	-0.18	-0.23

Note. *C*: feature-overlap parameter, *r*: rate parameter, σ : noise parameter, task: task effect, MD: effect of memory demand, DiffDev: group effect of intelligence-matched control group vs. general learning-difficulty group. The table contains estimates for the parameter means (Mean) and their standard deviations (SD) together with the T values (p-values are given in brackets). In the right part of the table (column six to nine) estimates for the parameter correlations across participants are given. Estimates are based on 1872 data points.

Table 6: Results of the IM for the group differences in parameter profiles.

Effect	Task	VMU			SMU		
	Contrast/ Parameter	overlap	noise	rate	overlap	noise	rate
Age	CG5/6 vs. CG2	0	+	+	0	+	+
DYS	CG2 vs. DYS2	-	+	0	-	+	0
GLD	CG5/6 vs. GLD5/6	0	0	+	0	0	+
	CG2 vs. GLD5/6	0	0	-	0	0	-

Note. A + represents more efficient performance for the respective first group, a - represents less efficient performance for the respective first group, 0 represents no difference. DYS: dyslexia, GLD: general learning difficulty, VMU: verbal memory-updating task, SMU: spatial memory-updating task, CG2: Control group second grade, DYS2: dyslexic group second grade, CG5/6: control group fifth/ sixth grade, GLD5/6: general learning-difficulty group fifth/ sixth grade.

Figures

Figure 1. Task examples of the verbal (top) and the spatial (bottom) memory-updating task.

Both task examples include MD-2. The example of the verbal task demands three updates and the spatial task four updates.

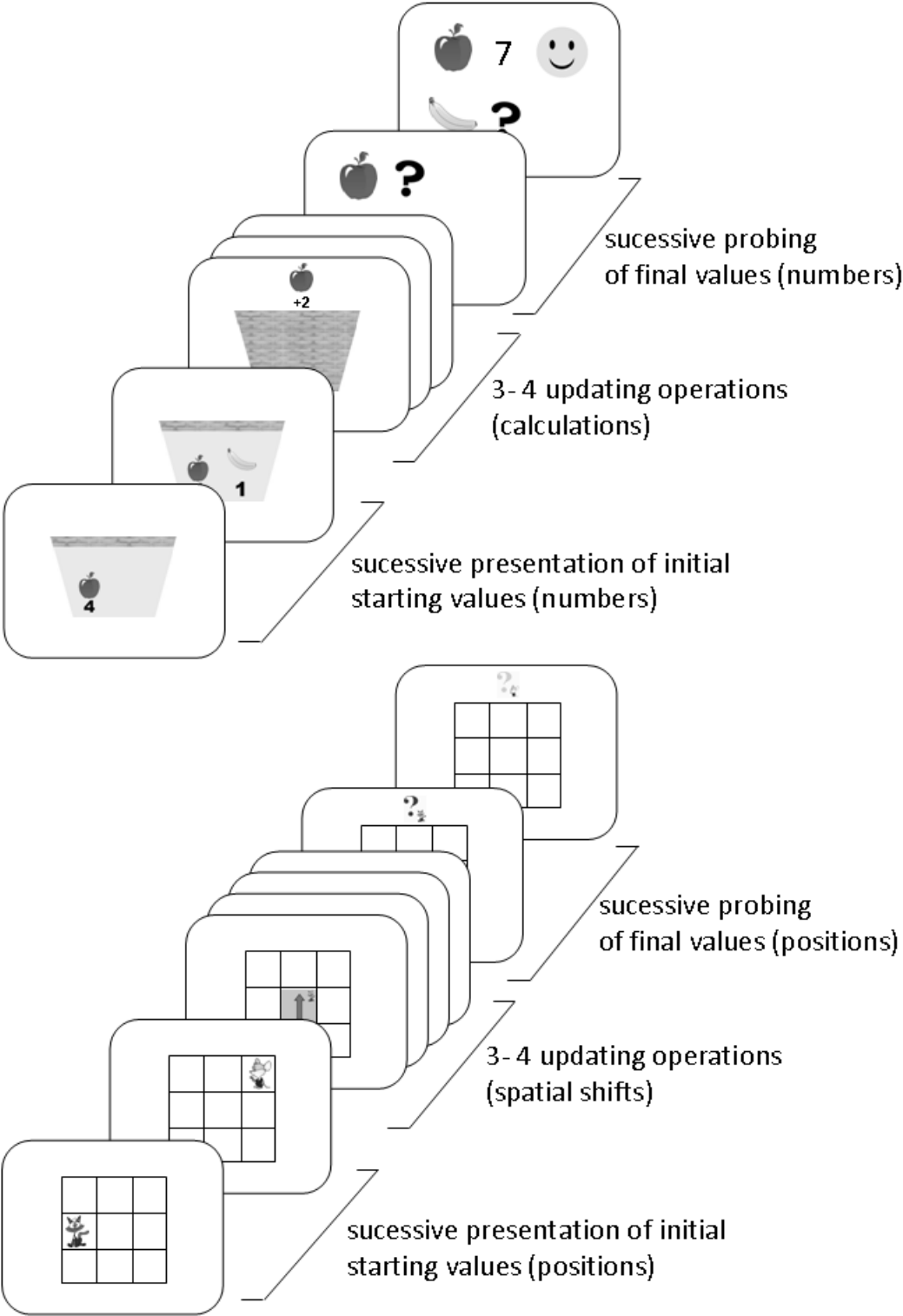


Figure 2. Data for the verbal (upper row) and the spatial (lower row) memory-updating task as a function of participant group, PT and MD (circle: MD-1 and triangle: MD-2) together with the predictions (lines) of the best-fitting IM.

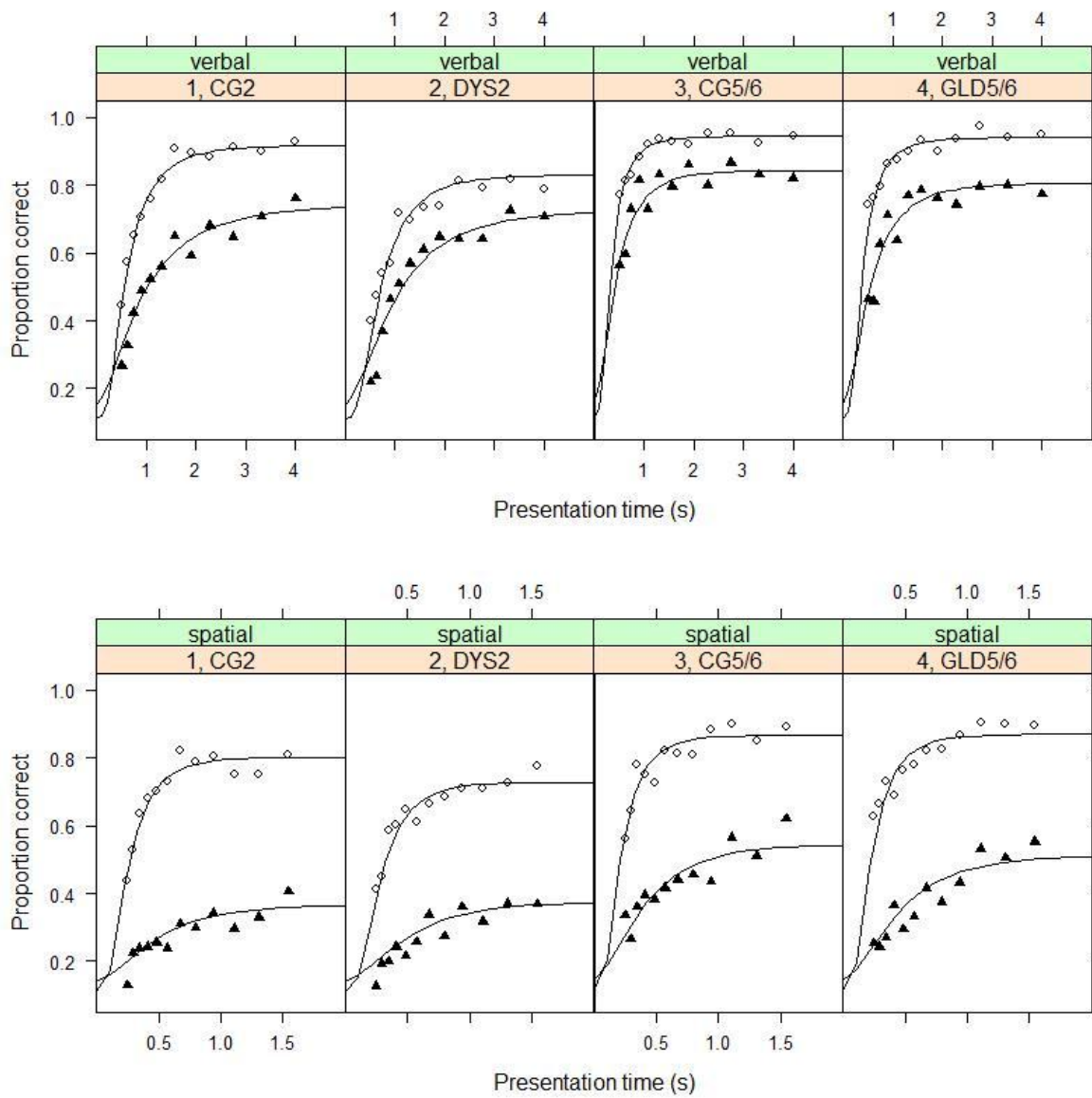
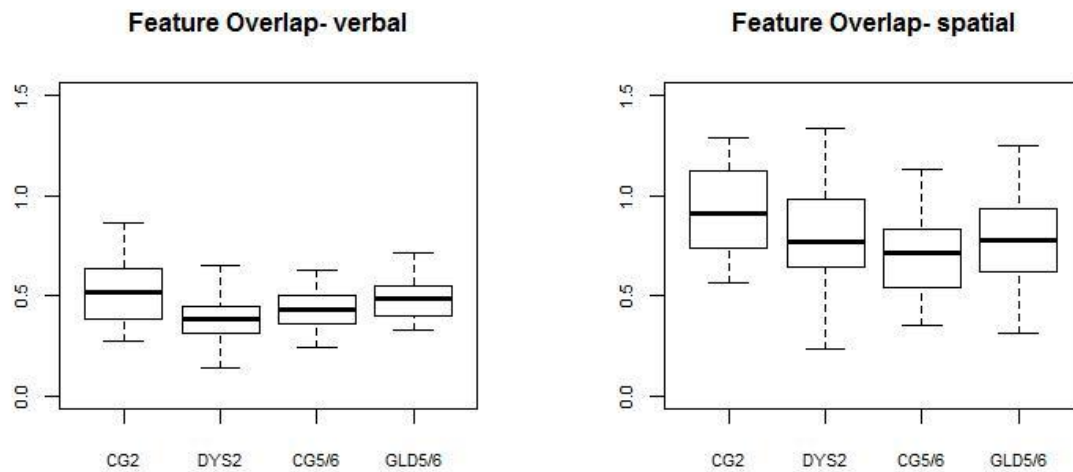
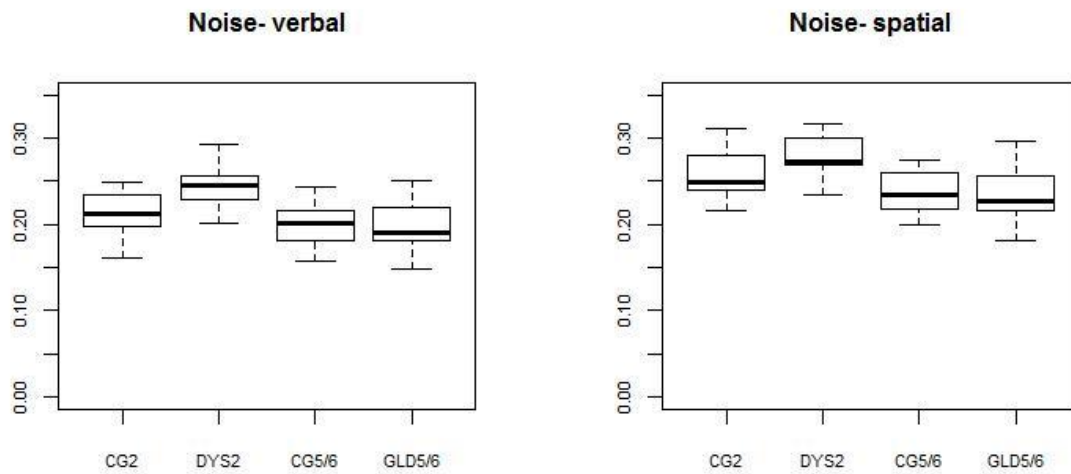


Figure 3. Boxplots of conditional modes of feature-overlap parameter (version 20) for the verbal (left) and the spatial (left) task as a function of group.



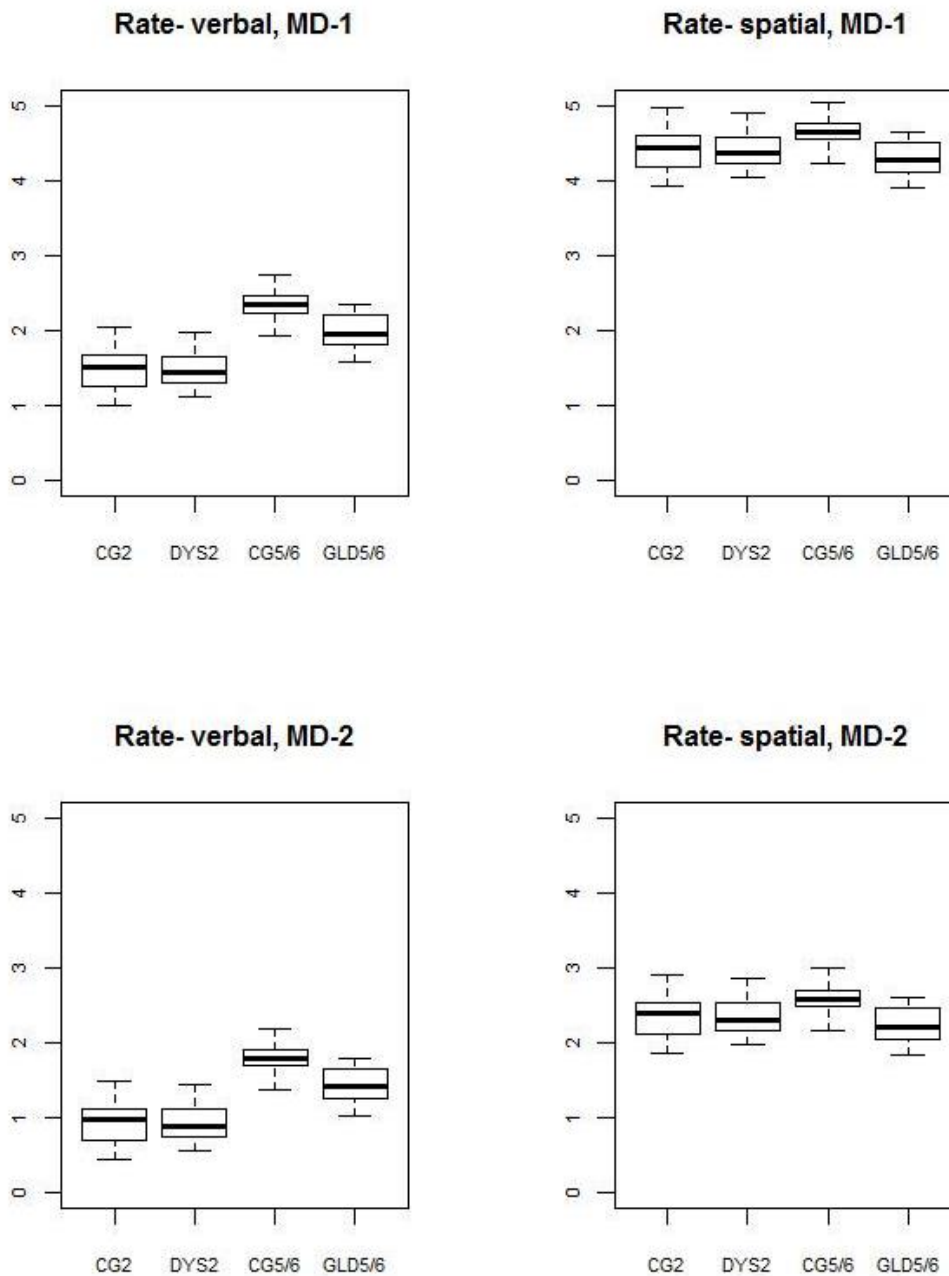
Note. The edges of the solid box correspond to the upper and lower quartiles of the respective distribution. The thicker line inside displays the median. The two whiskers show the overall range of the data, that is, minimum and maximum. CG2: Control group second grade, DYS2: dyslexic group second grade, CG5/6: control group fifth/ sixth grade, GLD5/6: general learning-difficulty group fifth/ sixth grade.

Figure 4. Boxplots of conditional modes of noise parameter (version 20) for the verbal (left) and the spatial (left) task as a function of group.



Note. The edges of the solid box correspond to the upper and lower quartiles of the respective distribution. The thicker line inside displays the median. The two whiskers show the overall range of the data, that is, minimum and maximum. CG2: Control group second grade, DYS2: dyslexic group second grade, CG5/6: control group fifth/ sixth grade, GLD5/6: general learning-difficulty group fifth/ sixth grade.

Figure 5. Boxplots of conditional modes of rate parameter (version 20) for the verbal (left) and the spatial (left) task as a function of group. Upper panels show the respective rate parameters for the MD-1 and the lower panels for the MD-2 condition.



Note. The edges of the solid box correspond to the upper and lower quartiles of the respective distribution. The thicker line inside displays the median. The two whiskers show the overall range of the data, that is, minimum and maximum. CG2: Control group second grade, DYS2: dyslexic group second grade, CG5/6: control group fifth/ sixth grade, GLD5/6: general learning-difficulty group fifth/ sixth grade.

Figure 6. Dotplots for 40 subjects of the two groups with older children (CG5/6 and GLD5/6). Children are ordered according to (a) individual scores of the figural STM task. Panels (b) to (d) display conditional modes of the overlap parameter of the verbal task, of the noise parameter of the verbal task, and of the rate parameter of the verbal task, respectively.

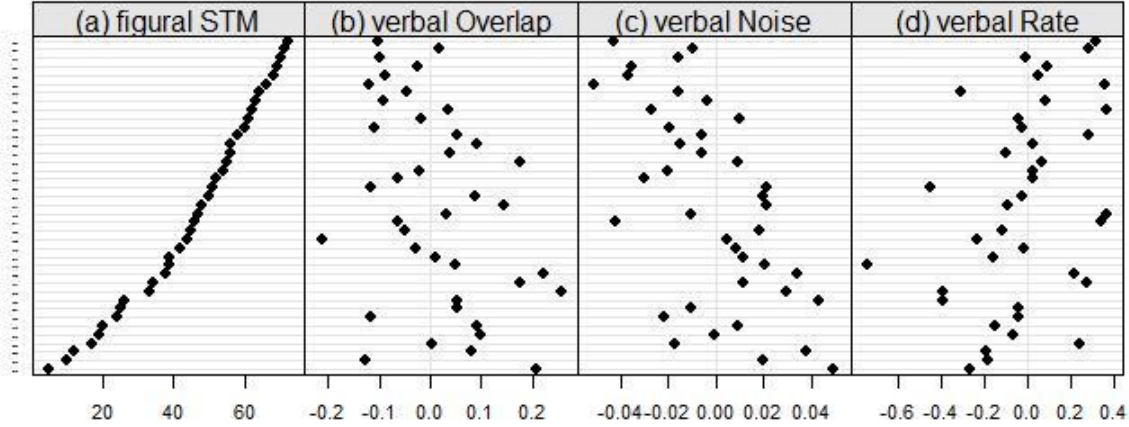
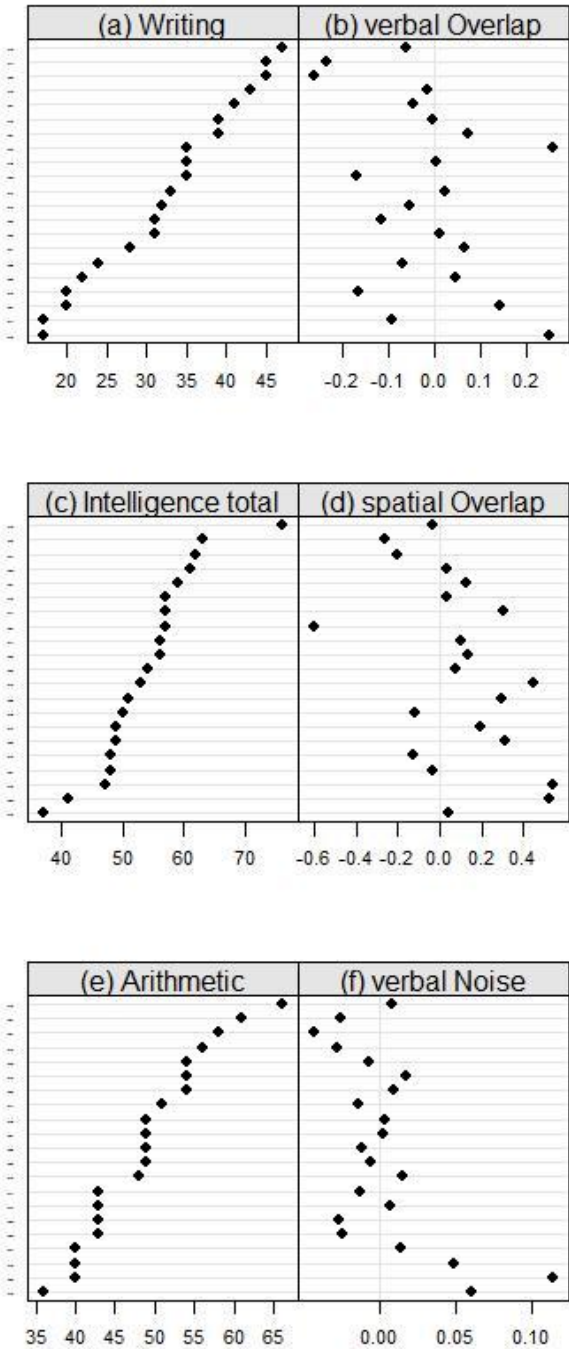


Figure 7. Dotplots for 20 subjects of the DYS2 group. In the upper part children are ordered according to (a) individual scores of the writing test. Panel (b) displays the conditional modes of the overlap parameter of the verbal task. In the middle part dyslexic children are ordered according to (c) their total intelligence scores and panel (d) shows the conditional modes of the spatial overlap parameter. In the lower part children are ordered according to (e) individual arithmetic test scores and panel (f) displays conditional modes of the noise parameter of the verbal task.



Appendix A

Interference model.

For a detailed description see Oberauer and Kliegl (2006). The expected proportion of features that is preserved for the representation of an item ($Prop_i$) in WM (i.e., is not lost through overwriting) is expressed as:

$$Prop_i / A_i = (1 - C/2)^{n-1} \quad (1)$$

where C is the mean proportion of features shared between any two items (i.e., feature overlap), and n is the number of elements held simultaneously in WM (i.e., MD). The function describes that with increased MD less features remain for each element. The proportion of preserved features for the representation of an item directly translates in its activation, A_i , which is therefore also given by Equation 1.

In order to retrieve an item from WM, its activation has to be transferred from the feature layer to its representation in the focus. This retrieval happens in a gradual fashion. It can be described by a negatively accelerated function:

$$a_i = (1 - C/2)^{n-1} (1 - \exp(-tr)) \quad (2)$$

where a_i is the activation of the target item in the focus layer, and the activation of that item in the feature layer, given by Eq. 1 acts as an asymptote of a_i , t is the time since beginning of the retrieval process (time for one updating cycle in our experiment, i.e., the PT), and r is the rate of activation. The gradual activation of the target item i by rate r represents the joint process of retrieving the target item into the focus of attention, and of activating the result of the updating operation.

However, not only the target item but also other items in WM receive activation. This is due to the partial feature overlap of all items in WM. Each competitor receives activation by those features it shares with the target. In general, each competitor will have grabbed away $C/2$ of the C feature units it shares with the target. Of the remaining $C/2$, other competitors (if present) are expected to grab away a proportion of $C/2$. The proportion of feature units of the

target that are shared with any competitor and remain bound to the target can, therefore, be expressed as $C/2$ times $(1-C/2)^{(n-2)}$ for $n > 1$. The activation equation for competitor items is therefore given by:

$$a_j = (C/2)(1 - C/2)^{n-1}(1 - \exp(-tr)) \quad (3)$$

The Boltzmann equation (Anderson & Lebiere, 1998, p. 80) describes the probability that among n items in WM the target item i forwards the highest activation to the focus layer:

$$p_i = \frac{\exp(a_i/T)}{\exp(a_i/T) + (n - 1) \exp(a_j/T) + (9 - n) \exp(0/T)} \quad (4)$$

whereby, p_i is the probability that the activation of a target item i , expressed as a_i , is higher than the activation of all other items transmitted to the focus layer: the target item with activation a_i , the other items currently in working memory, which have activation a_j , and the remaining items in the set which have activation 0. Parameter T reflects the standard deviation of activation, that is, the noise in the system. It is expressed by:

$$T = \sqrt{6} \sigma / \pi \quad (5)$$

The probability of recalling the correct item at the end of a trial, P_i , depends on the success of each individual updating step. For simplicity the model assumes independence of the probabilities for these succeeding updating steps. The probability to recall an item correctly at the end of a trial, therefore, is the product of the probabilities of successful individual updating operations on that item, times the probability that retrieval succeeds in response to the item's final probing. Therefore, accuracy in recalling each item i is given by:

$$P_i = 1/9 + (1 - 1/9)p_i^m p'_i \quad (6)$$

With chance level set to $1/9$ (since the participants were forced to select one of nine response alternatives -either spatial positions or numbers), m expresses the number of updating operations applied to item i , p_i the probability of a single successful updating step. The probability to succeed in the final retrieval, denoted p'_i , is computed just like p_i , but with processing time t set to infinity, because there was no time limit for retrieval.

Appendix B

Description of psychometric tests

For the psychometric tests that were applied in or prior to our experiment (subtests of the BUEGA test battery and memory tests) cronbachs alpha as a measure of reliability are given in brackets if available.

Verbal intelligence (analogies, .90): The experimenter reads up to 52 unfinished sentences in the form “A is X and B is ...” to the child. The child has to complete the sentence from memory. For some sentences multiple answers are defined as correct. A correct answer is scored with one point. Four wrong answers in succession lead to the stop of the test. The sum of the correct answers is scored.

Nonverbal intelligence (matrices, .93): The child sees three pictures in a two by two matrix. A fourth picture has to be found out of five different alternatives. The child has to point to one he/ she thinks belongs to the other three without necessarily specifying the rule. Up to 38 matrices are shown. A correct answer is scored with one point. Four wrong answers in succession lead to termination. The sum of the correct answers is scored. Nonverbal and verbal intelligence scores go down as intelligence total.

Expressive language (grammar, .93): The experimenter reads two sentences to the child. The second sentence is unfinished. The child has to finish the second sentence, which means to find the right grammatical form of a word used in the first sentence. Additionally a card with two pictures (one for each sentence) is shown, which should help to find the correct word. For example, the child has to finish: “This is one eye. These are two ...” with the word eyes. There are up to 57 items. A correct answer is scored with one point. For some items several answers are defined as correct answers. Six wrong answers in succession lead to termination. The sum of the correct answers is scored.

Reading (.94): The child has to read two cards with words. Card one includes 32 easy, card 2, 24 more difficult words. The experimenter stops the time for reading one card. During

reading the experimenter also marks errors if produced. Reading is terminated if card one is not finished after three minutes or if card two is not finished after six and a half minutes.

Reading time is scored. Reading errors (omissions, substitutions of phonemes, parts of words or whole words, complete distortion of words, separated reading of “sch”, “ng”, “nk”, “st”, “sp”) are scored. Reading time and reading errors go down as reading total.

Orthography (.82): The experimenter reads one word to the child. The child has to write it down. 16 words have to be written. Writing errors on the level of graphemes and phonemes are evaluated.

Calculating (.95): The experimenter reads one of up to 36 calculating tasks to the child. The child has to answer. No means can be used. A correct answer is scored with one point. Four wrong answers in succession lead to termination.

Attention (bp test, .87). The child has to mark the letters “b” and “p” in twelve rows of letters including also: d, g, q, h. Each row includes 50 letters. The number of targets in each list varies across rows. The child has four minutes to work through all rows. Errors and hits are scored up to the last letter that was marked by the child. Several measures of attention can be calculated. We report the percent of errors, that is, misses and false marks. Therefore the number of errors is divided by the number of hits and multiplied with 100.

Word recall forward. In the word recall forward test the experimenter speaks a list of one syllable words with one word every 1.5 seconds. The child has to recall the list of words. The number of word in each list increases across trials up to eight. Four trials are presented at each list length. The list length increases when the child responds correctly on at least two trials at a particular list length. Testing is discontinued when three trials of the same length are erroneously recalled. The maximum list length where at least two lists were recalled correctly is reported.

Digit recall backward. The experimenter speaks a list of digits with one digit per second. The child is asked to recall the sequence of digits in the reverse order. The number of

digits in each list increases across trials up to eight. Two trials are presented at each list length. The list length increases when the child responds correctly on at least one trial at a particular list length. Testing is discontinued when both trials of the same length are erroneously recalled. The maximum list length where at least one list was recalled correctly is reported.

Corsi-block (simple and complex): The child has to reproduce a sequence in the correct order by tapping on blocks irregularly positioned within a three dimensional array. Maximum sequence length is eight blocks. Four trials per sequence length are given. If the child responds correctly on at least two trials at a particular sequence length, number of sequences increases. Testing is discontinued when three trials of the same length are erroneously tapped. The maximum list length is reported where at least two sequences were recalled correctly. There is one test with simpler and one with more complex sequences.

Figural short-term memory-task (administered in our study). In the figural short-term memory-task (Oberauer, 1993) dots appeared in the cells of a ten by ten matrix sequentially for one second each. Participants had to recall the dot pattern. Responses were given by placing a cross in the corresponding cells in an empty matrix in the answer sheet. The score obtained for a response depended not on the absolute placement of the crosses, but only on their relative positions. Therefore, the relations between the dots and not their absolute positions in the matrix had to be remembered. After two practice items, participants worked on 15 items of this task. The number of dots increased by one every three items, ranging from two to six. The 15 testing items were identical for each participant.

Appendix C

Model-testing sequence for the Interference model

We checked goodness of fit of the different versions of the IM with log Likelihood Ratio Tests, Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The AIC and the BIC are based on the log Likelihood, taking into account the number of free parameters (AIC) and, in addition, the number of observations (BIC). Finally, we also report an adjusted R^2 statistic for each model (McElree & Doshier, 1989), that is the proportion of the observed variance predicted by the model adjusted by the number of free parameters. It is given by:

$$R_{adj}^2 = 1 - \frac{\sum_{i=1}^n (d_i - \hat{d}_i)^2 / (n - k)}{\sum (d_i - \bar{d})^2 / (n - 1)} \quad (1)$$

where d_i represents the observed values, \hat{d}_i are the predicted values, \bar{d} is the mean, n is the number of data points, and k indicates the number of free parameters.

No.	Fixed effects	Random effects	Df	AIC	BIC	Log-Lik	R_{adj}^2	Sign.
1	C, r, σ	C	5	-1028	-997	519.0	.511	-
2	$C + Task, r, \sigma$	C	6	-1671	-1633	841.4	.589	1
3	$C + Task, r, \sigma$	$C + Task$	7	-1790	-1746	902.1	.617	2

4	$C, r+ Task, \sigma$	r	6	-1252	-1214	632.0	.550	1
5	$C, r+ Task+ MD+ Task* MD, \sigma$	r	8	-2590	-2540	1303.0	.683	4
6	$C, r+ Task+ MD+ Task* MD, \sigma$	$r+ MD$	9	-2697	-2641	1357.7	.706	5
7	C, r, σ	σ	5	-1714	-1682	861.8	.598	-
8	$C, r, \sigma + Task$	σ	6	-1870	-1833	941.1	.615	7
9	$C, r, \sigma+ Task+ MD+ Task* MD$	σ	8	-2497	-2447	1256.4	.674	8
10	$C, r, \sigma+ Task+ MD+ Task* MD$	$\sigma + Task$	9	-2704	-2648	1361.1	.709	9
11	$C+ Task, r+ Task+ MD+ Task* MD, \sigma+ Task+ MD+ Task* MD$	$C+ Task, r+ MD, \sigma + Task$	17	-3492	-3386	1763.0	.791	-
12	$C+ Task, r+ Task+ MD+ Task* MD, \sigma + Task$	$C+ Task, r+ MD, \sigma + Task$	15	-3603	-3509	1816.5	.794	11
13	$C+ Task, r+ Task+ MD+ Task* MD, \sigma + Task$	$C+ Task, r, \sigma + Task$	14	-3605	-3518	1816.5	.795	11
14	$C+ Task + Age+ Dys, r+ Task+ MD+ Task* MD, \sigma + Task$	$C+ Task, r, \sigma + Task$	16	-3611	-3510	1821.3	.794	13
15	$C+ Task + Age+ Dys, r+ Task+ MD+ Age + Task* MD, \sigma + Task$	$C+ Task, r, \sigma + Task$	17	-3635	-3529	1834.6	.794	14
16	$C+ Task + Age+ Dys, r+ Task+ MD+ Age + Task* MD+ Task* Age, \sigma + Task$	$C+ Task, r, \sigma + Task$	18	-3641	-3529	1838.7	.795	15

17	<i>C+ Task + Age+ Dys, r+ Task+ MD+ Age + GLD+ Task*</i> <i>MD+ Task* Age, σ + Task</i>	<i>C+ Task, r, σ + Task</i>	19	-3645	-3526	1841.3	.795	16
18	<i>C+ Task + Age+ Dys, r+ Task+ MD+ Age + GLD+ Task*</i> <i>MD+ Task* Age, σ + Task+ Age</i>	<i>C+ Task, r, σ + Task</i>	20	-3659	-3534	1849.6	.795	17
19	<i>C+ Task + Age+ Dys, r+ Task+ MD+ Age + GLD+ Task*</i> <i>MD+ Task* Age, σ + Task+ Age+ Dys</i>	<i>C+ Task, r, σ + Task</i>	21	-3669	-3537	1855.4	.795	18
20	<i>19</i>	<i>19 with correlations</i>	31	-3687	-3493	1874.3	.794	19

Legend. *C*: feature-overlap parameter, *r*: rate parameter, σ : noise parameter, *task*: task effect, *MD*: effect of memory demand, *Age*: effect of age, *Dys*: effect of dyslexia (nested under younger group), *GLD*: effect of general learning difficulty (nested under older age group), *Df*: degrees of freedom, *AIC*: Akaike Information Criterion, *BIC*: Bayesian Information Criterion, *log-Lik*: log- Likelihood; R^2_{adj} : adjusted R^2 statistic, *Sign.*: model fit is significantly better than the model with the number in this column.

Appendix D

Model-testing sequence for the Interference model for the GLD5/6- CG2 comparison

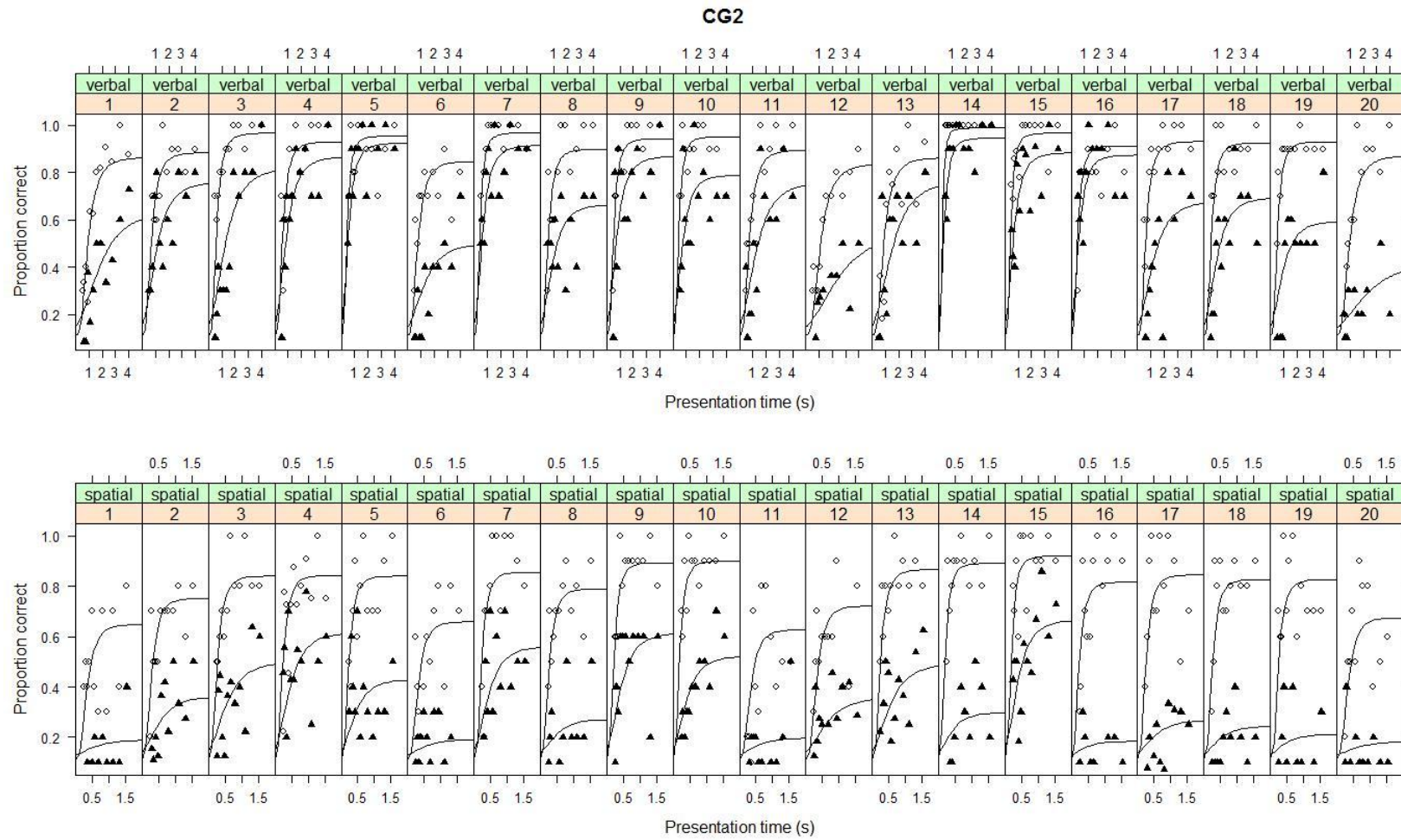
No.	Fixed effects	Random effects	Df	AIC	BIC	Log-Lik	R ² _{adj}	Sign.
1	C, r, σ	C	5	-650	-622	330.1	.567	-
2	$C+ Task, r, \sigma$	C	6	-991	-957	501.3	.642	1
3	$C+ Task, r, \sigma$	$C+ Task$	7	-1043	-1005	5287.1	.665	2
4	C, r, σ	r	5	-230	-202	119.9	.416	-
5	$C, r+ Task, \sigma$	r	6	-260	-227	136.0	.425	4
6	$C, r+ Task+ MD, \sigma$	r	7	-1180	-1141	596.8	.682	5
7	$C, r+ Task+ MD, \sigma$	$r+ MD$	8	-1234	-1190	625.1	.710	6
8	$C, r+ Task+ MD+ Task* MD, \sigma$	$r+ MD$	9	-1456	-1406	737.0	.726	7
9	C, r, σ	σ	5	-946	-918	477.9	.637	-
10	$C, r, \sigma + Task$	σ	6	-1026	-993	518.9	.652	9
11	$C, r, \sigma+ Task$	$\sigma + Task$	7	-1092	-1053	553.0	.683	10
12	$C+ Task, r+ Task+ MD+ Task* MD, \sigma+ Task$	$C, r, \sigma + Task$	13	-1833	-1761	929.4	.804	-
13	$C+ Task, r+ Task+ MD+ DiffDev+ Task* MD, \sigma + Task$	$C, r, \sigma + Task$	14	-1838	-1761	933.2	.804	12

14	$C + Task, r + Task + MD + DiffDev + Task * MD, \sigma + Task$	$C, r, \sigma + Task$	15	-1842	-1758	935.8	.804	13
15	14	14+ correlations	21	-1844	-1728	942.9	.804	14

Legend. C : feature-overlap parameter, r : rate parameter, σ : noise parameter, task: task effect, MD: effect of memory demand, DiffDev: group effect of intelligence-matched control group vs. general learning-difficulty group. Df: degrees of freedom, AIC: Akaike Information Criterion, BIC: Bayesian Information Criterion, log-Lik: log- Likelihood; R^2_{adj} : adjusted R^2 statistic, Sign.: model fit is significantly better than the model with the number in this column.

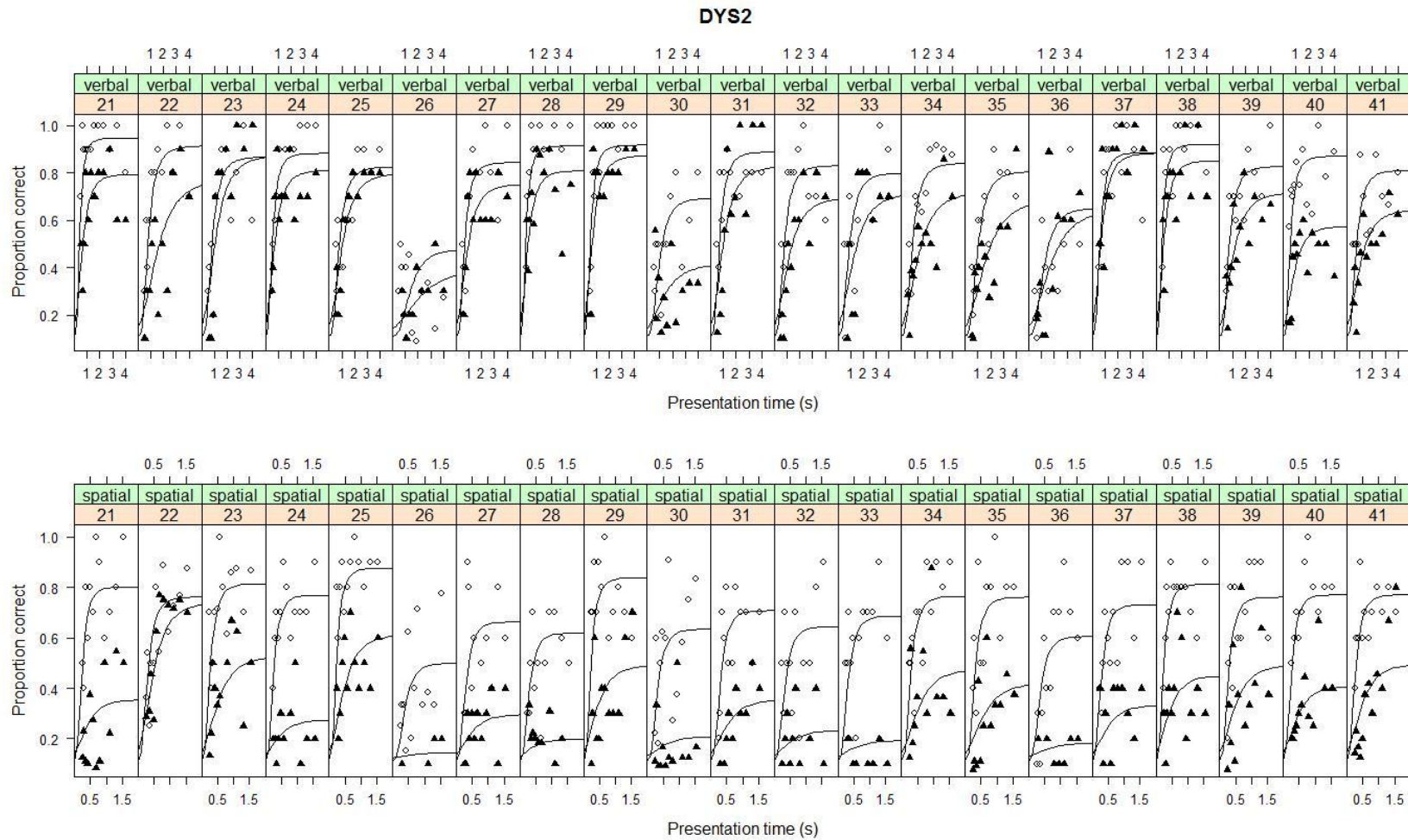
Appendix E

Individual data for the verbal (upper row) and the spatial (lower row) memory-updating task for the participants of the CG2 group as a function of PT and MD (circle: MD-1 and triangle: MD-2) together with the predictions (lines) of the best-fitting IM



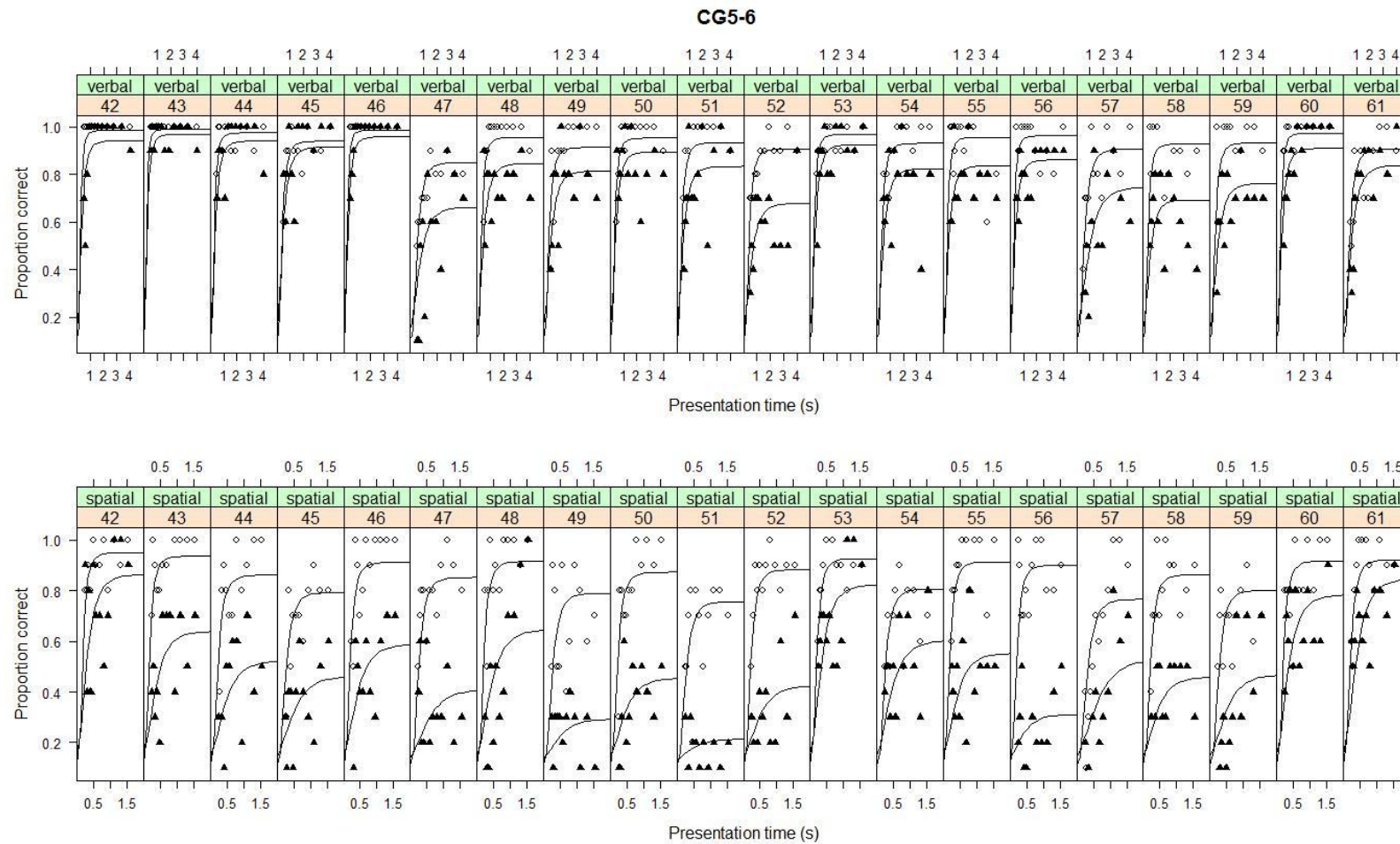
Appendix F

Individual data for the verbal (upper row) and the spatial (lower row) memory-updating task for the participants of the DYS2 group as a function of PT and MD (circle: MD-1 and triangle: MD-2) together with the predictions (lines) of the best-fitting IM



Appendix G

Individual data for the verbal (upper row) and the spatial (lower row) memory-updating task for the participants of the CG5/6 group as a function of PT and MD (circle: MD-1 and triangle: MD-2) together with the predictions (lines) of the best-fitting IM



Appendix H

Individual data for the verbal (upper row) and the spatial (lower row) memory-updating task for the participants of the GLD5/6 group as a function of PT and MD (circle: MD-1 and triangle: MD-2) together with the predictions (lines) of the best-fitting IM

