Development of reading and phonological skills of children at family risk for dyslexia: A longitudinal analysis from kindergarten to sixth grade

Sophie Dandache, Jan Wouters, & Pol Ghesquière

Introduction

Developmental dyslexia is a highly hereditary learning disability characterized by severe reading and spelling difficulties (Gersons-Wolfensberger & Ruijssenaars, 1997). These difficulties are persistent and resistant to usual teaching methods and remedial efforts. The prevalence of dyslexia lies between 5% and 10% (Maughan, 1995). Providing a clear definition of dyslexia is important to help distinguishing between reading difficulties that are caused by poor instruction in reading, limited exposure, or poor schooling from those that are due to cognitive and biological factors. One of the main differences between these two is that solely the latter implies a predisposition from birth hence their difficulties cannot be remediated by didactical measures and remedial efforts (Vandewalle, Boets, Ghesquière, & Zink, 2010).

The phonological deficit in dyslexia

Learning to read requires the acquisition of phonological skills. A deficit in these skills would affect the grapheme-phoneme decoding which is the first and most critical phase in learning to read (Wimmer & Schurz, 2010). Hence, poor phonological abilities would constitute the main origin of the problems encountered by dyslexic readers (Vellutino, Fletcher, Snowling, & Scanlon, 2004; Shaywitz & Shaywitz, 2003).

It is noteworthy that phonological ability is an umbrella term used to describe the ability to access, process, and manipulate speech sounds (Wagner & Torgesen, 1987). The phonological deficit becomes manifest in three interrelated broad areas: *phonological awareness (PA)*, *rapid automatized naming (RAN)* and *verbal short-term memory (VSTM)*

(Ramus & Szenkovits, 2008; Vandewalle et al., 2010). In the next section we will present a short description of each of these three areas.

Phonological awareness, refers to the conceptual understanding and explicit awareness that spoken words consist of individual speech sounds (phonemes) and combinations of speech sounds (syllables, onset-rime units) (Lyon, Shaywitz, & Shaywitz, 2003; Torgesen, Wagner, Rashotte, Burgess, & Hecht, 1997). Furthermore, PA involves three skills that could be simultaneously affected in dyslexia: (a) identifying and differentiating between letters, (b) processing phonological information, and (c) relating specific letters to specific sounds (Froyen, Willems, & Blomert, 2011). Dyslexic readers have more troubles processing speech sounds hence their speech perception is poorer and less precise (Griffiths & Snowling, 2002).

Naming speed is defined as the efficiency of phonological code retrieval from long-term memory and is an influential process of reading (Vellutino et al., 1996; Denckla & Rudel, 1976; Wolf, Bally, & Morris, 1986; Denckla & Cutting, 1999). Authors claim that naming speed deficits would be due to a pervasive problem in the underlying phonological representations of words (Snowling, Bishop, & Stothard, 2000; Savage, Pillay, & Melidona, 2008). Furthermore, an independent contribution of naming speed to reading acquisition has widely been observed (Wolf & Bowers, 2000; Kirby, Silvestri, Allingham, Parrila, & La Fave, 2008; Catts, Gillispie, Leonard, Kail, & Miller, 2002). Naming speed ability is typically assessed through the RAN procedure. Several studies showed that in transparent orthographies such as German or Finnish, RAN repeatedly showed to have a strong predictive link to reading (Eklund, Torppa & Lyytinen, 2013; Holopainen, Ahonen, & Lyytinen, 2001; Wimmer & Mayringer, 2002). Additionally, dyslexic readers have been proved to be significantly slower in these tasks, confirming the existence of a link between RAN and reading deficits (Wolf, Bowers, & Biddle, 2000).

Finally, reading acquisition requires the recognition of an array of letters, words and their meaning (Vellutino, Fletcher, Snowling, & Scanlon, 2004). Learning to read is then linked to the development of VSTM which is the system responsible for the storage, retrieval and processing of stable associations between spoken and written material (Mann & Liberman, 1984; Kibby, 2009; Swanson, Kehler, & Jerman, 2010; Savage &

Frederickson, 2006). Several studies have demonstrated impaired VSTM performance within dyslexic children and adults (Chiappe, Hasher, & Siegel, 2000; Jeffries & Everatt, 2004). These difficulties are however more apparent in beginning readers who are constantly exposed to new reading material (Katz, Shankweiler, & Liberman, 1981; Kibby & Cohen, 2008; Kibby, Marks, Morgan, & Long, 2004). At a later stage, when children have developed their semantic knowledge, they can start relying on the context in order to compensate for memory deficits (Snowling, 2000; Kibby & Cohen, 2008). It is nonetheless notable that some studies failed finding a deficit in VSTM's performance of dyslexic readers in comparison to age-matched controls (Bowers, 1995). Kipp and Mohr (2008) argue that these findings could be due to the use of different definitions for dyslexia.

Exploring the evolution of reading with age

A main interest in analyzing developmental dyslexia is to explore what skills, prior to the onset of formal reading instruction, are the best predictors of later literacy ability. Finding these skills helps the orientation and implementation of early reading treatment which is of high importance to obtain effective results (Peterson & Pennington, 2012). An increasing number of longitudinal studies have been conducted to investigate reading and cognitive skills of normal as well as dyslexic readers.

A review on prediction studies conducted in 1998 by Scarborough (in Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004) concluded that the best kindergarten predictors of later reading skills were letter knowledge followed by PA, RAN, verbal memory tasks, intelligence and vocabulary. It is however noteworthy that predictive validity of measures collected in preschool might vary in function of the nature of the sample, the orthographic transparency of the language (i.e., the level of correspondence between the written symbols and their sounds), the length of the follow-up and the nature of the outcome (Schatschneider et al., 2004; Ziegler & Goswami, 2005). Family risk is however the strongest predictor for dyslexia with a 33 to 66% chance for a child coming from a high-risk family to develop dyslexia (Van Bergen, 2013). With regards to cognitive skills, although PA has repeatedly proved to be a strong predictor for later reading abilities, research investigating its prediction power present a consistent amount of variability. An important moderator that if often mentioned in explaining the differential importance of precursors of later reading abilities is orthographic transparency (de Jong & van der Leij, 2003; Puolakanaho et al., 2008). For instance, Van Bergen (2013) and Blomert and Willems (2010) found no group differences in kindergarten on PA .

With regards to cognitive skills, PA has repeatedly been shown to be the strongest predictor of literacy development in opaque orthographies as English (Wagner, Torgesen, & Rashotte, 1994; Caravolas, Hulme, & Snowling, 2001). In more transparent orthographies as in Dutch or Finish PA's prediction power seems to be weaker (Caravolas et al., 2012; Puolakanho et al., 2008; Van Berge, 2013; Wimmer, Mayringer, & Landerl, 2000). In a cross-linguistic meta-analyses Landerl, Ramus, Moll and Lyytinen (2013) compared six European languages that they classified in three degrees of orthographic complexity: opaque (French and English), medium (Dutch and German) and transparent (Hungarian and Finish). Results showed that the more opaque the orthography, the strongest was PA's prediction power. These results are at odds with our study conducted in transparent Dutch where significant differences in PA were found already in kindergarten (Boets et al., 2003; 2007; 2010).

There exists however discrepancy in the literature concerning the relation between RAN and reading. For instance, a cross-linguistic longitudinal study comparing four languages (English, Spanish, Slovak and Czech) showed that, in the four languages, RAN measured before the onset of reading was as important in predicting later reading and spelling as PA (Caravolas et al., 2012). In transparent languages as Dutch, PA tends to be more important in the initial phases of reading development while RAN seems to gain importance in a later stage (Verhagen, Aarnoutse & Van Leeuwe, 2010; Vaessen & Bloemert, 2010). The same results were however found in the opaque English orthography where the relative contribution of RAN to reading increased with age (Kirby, Parrila & Pfeiffer, 2003). In contrast, other studies conducted on opaque orthographies showed that RAN was mostly linked to reading in the earlier grades and its influence diminished in time (Torgsen et al., 1997; Georgiou et al., 2008).

Finally, few studies have found a significant relation between VSTM and reading development. A longitudinal study on English-speaking children at high and low risk for dyslexia showed that, at 3 years and 9 months, the at-risk non-dyslexic children performed better than at-risk dyslexic readers but poorer than low-risk non dyslexic readers on non-word repetition (Snowling, Gallagher & Frith, 2003; Snowling, Muter & Caroll, 2007). Boets et al. (2007, 2010) found a significant predictive value for reading accuracy in first grade these results were however not confirmed by other similar studies (de Bree , Wines & Gerrits , 2010; Blomert , 2010). De Jong and van der Leij (1999) explain this lack of effect by the fact that VSTM would intertwine with phonological awareness and recoding. Exploring reading growth has led to two distinct theories characterizing the reading developmental trajectories of normal readers and dyslexic readers. On the one hand, the 'lag model' stipulates that children with dyslexia start at a lower level than their peers but show a steeper evolution leading them to catch up with good readers over time (Stanovich, Nathan, & Zolman, 1988). On the other hand, defenders of the 'deficit model' argue that the gap between poor and good readers persists over time with a possibility for it to grow. The latter theory has been approved by studies demonstrating that good readers start with a higher intercept than dyslexic readers but follow a similar *quadratic* trajectory (Wei, Blackorby, & Schiller, 2011; Catts, Bridges, Little, & Tomblin, 2008). Additionally, the question whether reading growth follows a linear or a quadratic trajectory has also been subject of debate. On the one hand, several authors defended the hypothesis of a quadratic trajectory. This theory can easily be explained by the fact that decoding efficiency witnesses its steepest evolution in earlier grades when learning to read. Later, with the increase of reading experience, the reader's performance tends to stabilize leading to a decrease in growth rate hence a quadratic growth shape (Aarnoutse, van Leeuwen, Voeten, & Oud, 2001; Nakamoto, Lindsey, & Manis, 2006; Wei, Blackorby, & Schiller, 2011). On the other hand, Jordan, Kaplan and Hanich (2002) followed children from second to third grade reporting *linear* growth in reading. Yet, in the latter only data from two testing waves were used.

The present study

This research is an extension of a previous study (Boets et al., 2010) where Dutch-speaking children at high-risk of dyslexia and their matched controls, referred to as 'low-risk' children, have been selected in Flanders (Belgium) and tested on phonological and reading measures at three time points (kindergarten, first and third grade). Once the children’s reading performance could be well assessed (i.e., 3rd grade), Boets et al. (2010) divided this sample in three groups: dyslexic readers (DR), normal readers at high-risk for dyslexia (HR-NR) and normal readers at low-risk for dyslexia (LR-NR). In order to explore the further development of reading abilities, the same type of tasks where conducted again when children attended 6th grade. The addition of this measurement wave allows us to investigate the three research questions discussed here below.

Before we move on to our objectives, it is worth noting that the ‘at risk’ criterion refers to the family occurrence of dyslexia which is traditionally used (Gilger, Pennington, & DeFries, 1991). The reasons for selecting children at high-risk were twofold. On the one hand, the hereditary characteristic of dyslexia has been repeatedly confirmed by longitudinal or twins studies were the genetic factor has been approved showing that 30 to 50% of children at family risk of dyslexia ultimately developed reading difficulties (Francks, MacPhie, & Monaco, 2002). Hence, selecting a sample of high-risk children would help increasing the sample size of our ‘dyslexic’ group. On the other hand, a comparison of children at high and at low risk of dyslexia would give us an indication concerning the multi-componential nature of dyslexia. More precisely, it would allow us to test whether dyslexia is an all-or-none condition or rather a continuum where only some children would fill all the criteria for dyslexia (Snowling, Gallagher, Frith, 2003).

This leads us to our first objective. The first analyzes conducted on our sample have confirmed the presence of a gap between HR-NR and LR-NR where the former performed significantly poorer on tasks that required sublexical phonological representation (i.e., spoonerism and non-word repetition). In addition, and probably as a consequence of this sublexical impairment, HR-NR showed significantly weaker scores on letter knowledge and spelling in first grade and pseudoword reading accuracy and spelling in third grade (Boets el al., 2010). These results confirm the theory stating that reading difficulties in dyslexia are not an all-or-none condition. However, the addition of a testing wave requires the modification of the categorization criteria for dyslexic readers (see methods) hence the repartition of the sample in each of the three groups. Our ***first objective*** is then to perform new group comparisons at a later time point to verify whether the differences observed between the three groups are still tenable. The addition of criteria for defining dyslexic readers is expected to result in a clearer distinction between the groups. Consequently, larger differences should be observed between the three reading groups on both phonological and reading scores.

The ***second objective*** of this study is to explore the developmental trajectories of reading and phonological processes through growth curve modeling. Given that in the earlier analyzes of our data, Boets et al. (2010) solely analyzed the evolutionary profile of data collected in first and third grade, it is not surprising that a linear profile was observed. However, as children have acquired a fair decoding level by third grade, we do expect a decrease in the evolution after the addition of the last wave hence the growing curve would no more be linear. Furthermore, following the deficit evolutionary model, we expect the three groups to differ in their average score while conserving a similar evolutionary profile.

Finally, the ***third objective*** is to assess the relation between phonological skills and initial reading status as well as their implication in explaining growth variation. We expect PA and RAN to have the highest impact on reading achievement in all groups confounded.

Method

As mentioned in the previous section, this study is an extension of an earlier one where measures of literacy and phonology were collected in Flanders (Belgium) within Dutch-speaking children (Boets et al., 2010). In addition to the measures collected in kindergarten, first and third grade, the same sample was tested again in 6th grade which allows us to further investigate the development of reading skills. Initially, the sample comprised sixty-two children (36 boys and 26 girls) at high (HR) and low risk (LR) of dyslexia. On the one hand, half of the participants (N=31) constituted the HR group and had at least one first-degree relative with a formal diagnosis of dyslexia. On the other hand, the LR group (N=31) had no history of reading disabilities in the family. The latter group was matched with the HR on 5 criteria: (1) educational environment, i.e. same school, (2) sex, (3) age, (4) non-verbal intelligence, and (5) parental educational level (for further details see Boets et al., 2010). Additionally, the vocabulary subtest of the Wechsler Intelligence Scale for Children 3rd edition (WISC-III; Wechsler, 1992) were administered in first grade but the results were not taken into consideration for the children's selection. None of the participants had a history of brain damage, psychiatric disorder, hearing loss nor visual problem.

In a second step, by considering reading and spelling scores in first and third grade, each of the HR and LR groups was subdivided in two groups of dyslexic readers (HR-DR vs. LR-DR) and normal readers (HR-NR vs LR-NR). However, to take into account both the severity and the persistence of children's literacy problems (Gersons-Wolfenberger & Ruijssenaars,1997), we adjusted the inclusion criteria for the dyslexic group after the addition of 6th grade reading. Furthermore, a second criterion was added to exclude readers with solely a spelling difficulty. Hence, a participant has been classified as dyslexic if he/she (1) obtained a score lower than percentile 10 on at least two measurement moments for the spelling (Dudal, 1997) and/or reading task (van den Bos, Spelberg, Scheepstra, & de Vries, 1994) *and* (2) a score lower than percentile 50 on reading and spelling tasks on all testing moments. It is noteworthy that, for the last data collection, 5 participants have left the study. This final sample consisted of 29 HR children 10 of them diagnosed with dyslexia (10/29 = 34%) and 28 LR of which 3 developed dyslexia (3/28 = 11%). As was observed in the previous article, both dyslexic groups did not significantly differ on any of the administered tasks. Hence, both dyslexic groups were collapsed, leaving us with a sample of 13 DR, 19 HR-NR and 25 LR-NR.

Materials and Procedures

The phonological tests were selected in a way that reflected the three traditional domains of phonological processing: PA, RAN and VSTM and tests of literacy skills (van den Bos, Spelberg, Scheepstra & De Vries, 1994). The tasks used for testing kindergarten children were: *first-sound, end-sound, and rhyme identification, first-sound and rhyme categorization* and *rhyme production.* Starting from the 1st grade, the following tasks used were: *phoneme deletion* and *spoonerism.* While two RAN tasks were administered in kindergarten (i.e., *colors* and *objects*), two other ones were added for the higher grades (i.e., *letters* and *digits*). For all grades the same VSTM tasks (*Non-word repetition* and *Digit* span) were administered. Literacy skills were measured using *letter knowledge* for kindergarten and using the '*One-Minute-Reading test*' (a standardized word reading test)

(Brus & Voeten, 1973), the *Klepel (a standardized non-word reading test)* (Van den Bos et al., 1994) and a standardized spelling test (Dudal, 1997) for grade 1, 3 and 6.

In the next section, we will give a brief description of the different tasks, more details can be found in Boets et al. (2010).

Measuring PA

First-sound, end-sound, and rhyme identification. In these three tasks, the child was presented with 5 pictures and was asked to point at the word that contained the same (1) first sound, (2) end sound (10 items each) and (3) rhyme sound (12 items) as a given word between different distracters (de Jong, Seveke, & van Veen, 2000, adapted by van Otterloo and Regtvoort).

Rhyme production. The aim of this task is to test whether the child is able to produce a word that rhymed correctly with another one. The child was presented with a one-syllable word (8 items in total with increasing difficulty) and was asked to produce a rhyming word or non-word so they are not penalized by their vocabulary knowledge.

First-sound and rhyme categorization. For each of these two categorization tasks, the child was presented with three monosyllabic words from which he/she had to name the two that were similar with respect to the first sound (15 items) or the same rhyme (20 items).

Phoneme deletion. In this task, the examiner orally presented a pseudo-word and the child was instructed to provide a new word after deleting the initial or last phoneme. The task has two levels of difficulty: in the first part, the child is provided with a series of 10 pseudo-word that become a real word after deletion of the given phoneme. The second part (that includes 18 items) becomes more complicated since the pseudo-words remain meaningless after following the instructions.

Spoonerism. This task is divided into three parts with increasing level of difficulty, each one composed of two practice items and ten test items. In the first block, the participant is instructed to exchange the first phoneme of a real word with another phoneme provided by the instructor. In the first five items, the word changes meaning after exchanging the phoneme to become another word. In the second part, the participant is instructed to exchange the initial phonemes of two presented words for the first five items to create two new words (e.g., BAAS-VOL become VAAS-BOL) and two words to create two new pseudo-words for the next five items (e.g., DIKKE-BIL become BIKKE-DIL). Finally, the last part is similar to the second one, only this time the child is asked to exchange the first two consonants of the given words (e.g., KRANT-PLAS become PLANT-KRAS). Each correct answer was rewarded with one point for the first set and two points for the second and third set. The maximum score was then 50. While the first set had to be totally completed, the second sets was stopped after four consecutive errors (0/2 for an item). The child was assessed on the third set only if he/she scored equal or above 6/20 on the second set.

Measuring RAN

The test assesses the rapid serial naming for 5 familiar colors, objects, numbers, and letters (Van Den Bos, Zijlstra, & Lutje Spelberg, 2002). For each type of symbol, the child is presented with a card of 50 stimuli randomly arranged, with each stimulus appearing 10 times. The child is instructed to read the symbols as fast and accurately as possible.

Measuring VSTM

In the *Non-word Repetition Test (NRT)*, the child hears a non-word that he/she is instructed to repeat orally immediately after their presentation. The items are divided into four conditions of different length. Every category contains twelve test items with increasing number of syllables. The non-words are recorded on a computer by a professional speech-therapist and are presented to each child through a headphone at a fixed level of 70 dB SPL to the right ear.

Digit span (DS), consists in the immediate serial recall of spoken lists of digits between 1 and 9. The test comprises 21 items, divided in 7 sequences of numbers with a list length increasing from 2 to 9 digits, each list comprising 3 stimuli of the same length. Testing continues until the child fails on two of three trials of the same list length. The digits are recorded on a computer by a professional speech-therapist and are presented to each child through a headphone at a fixed level of 70 dB SPL to the right ear.

Literacy measures

Letter knowledge. In this task, the child was asked to name (or sound) the 16 most frequently used letters in Dutch books.

One Minute Test (Standardized word reading task). This standardized task combines speed and accuracy since the child is instructed to read in one minute as fast and accurately as possible a word list with increasing difficulty (Brus & Voeten, 1973).

*Klepel (Standardized pseudoword reading task).*The child must, within a time limit of two minutes, read a list of pseudo-words as fast and as accurate as possible (van den Bos, Spelberg, Scheepstra,& de Vries, 1994).

Spelling test. Children are asked to spell isolated single word, single words in a sentence context, and short sentences (Dudal, 2006). The maximum score on the test was 60 with grade-appropriate versions.

Statistical analysis

Statistical analysis of the data proceeded in several different steps. In the first section, we present descriptive statistics conducted on the phonological tasks first and then on the literacy ones. More precisely, we start by assessing the reliability and correlations of the different tasks. Next, we create factor scores for each grade separately by the means of a principal component analysis with varimax rotation. Finally, planned contrasts are used to compare the different scores between the three reading groups.

In a second step, a linear mixed model with individuals as random factors is fitted to the data. This model takes into account the non-independence of the reading scores within the same subject. Estimates for the fixed effects were used to test for differences in the average evolution of the phonological and reading scores in the three groups. Random intercepts and random slopes for time are included to model the covariance structure. In a third step, the three phonological scores are successively added to an unconditional model with the factor score for reading as a dependent variable in order to investigate their link to the initial reading status and its growth rate. All models were estimated with the default estimations methods in the SAS procedures PROC MIXED (for the linear mixed model) and PROC GLIMMIX (for the logistic mixed model). It is worth noting that one of the advantages of using PROC MIXED is that it does not penalize for unequal time intervals (Howell, 2010)

Descriptive statistics

Phonological abilities

Cronbach's alpha were calculated to assess the reliability of phonological measures (see table 1.a to 1.d). To examine the correlation of measurements within subjects, Pearson correlations between phonological factors assessed at different time points showed stability in all administered tests (all p's < 0.001).

Consistently with what has been applied in earlier grades (see Boets et al., 2003), principal component analysis with varimax rotation for orthogonal phonological factors were performed on the measures administered in 6th grade. Based on the eigenvalue rule of thumb (i.e., eigenvalue >1), three factors enclosing separate measures of RAN, PA and VSTM were extracted.

Since the children were distributed in new groups, we re-calculated the descriptive statistics for the earlier grades (Boets et al., 2010). Descriptive statistics on phonological measures are described in table 2. Factor scores were transformed to effect sizes relative to the mean and standard deviation of the LR-NR group to enhance the interpretation of the results. Results show that dyslexic readers score significantly lower than normal readers on PA and RAN in each testing wave. For VSTM, dyslexic readers score significantly lower than normal readers on NRT from kindergarten to third grade. No significant difference is ever found between the three groups for DS. As expected, the HR-NR group scores between the DR and LR-NR group. Finally, factor scores suggest that the phonological deficit for the group of dyslexic readers increased with time especially for PA.

Due to the strong theoretical evidence in dyslexia literature, planned contrasts were performed on all measures for each time point (see table 2). The dyslexic group scored significantly lower than the LR-NR on all PA, RAN and VSTM measures at all time points except for NRT in 3rd and 6th grade and digit span at all time points.

Finally, the HR-NR group performed at an intermediate level for all tasks at all time points. These differences were however not always significant. More precisely, compared to the dyslexic group, they scored significantly higher on all PA and RAN tasks from the 1st, until the 6th grade except for spoonerism in the 1st grade. However, in comparison to the LR-NR, they scored significantly lower only on NRT in kindergarten.

Literacy measures

Descriptive statistics were calculated for the scores of the three groups on the different literacy measures (table 3). Planned contrasts revealed that the dyslexic group scored significantly lower than the two non-dyslexic groups on all reading measures at all time points (see table 3). The HR-NR performed at an intermediate level for all tasks at all time points. These differences were however solely significant for non-word reading and spelling (which has already been observed in earlier analyzes) in 3rd grade and word-reading in 6th grade. Pearson correlations assessed at different time points showed a stability in the relative position of the participant's literacy measures (all p's $< .01$). Additionally, the strongest correlations were observed between measures of 3rd and 6th grade (see table 4). This observation is not surprising and can easily be explained: as children in Flanders start to read in the first grade, their reading abilities are still very limited. Later, when they reach third grade, two years of reading experience renders their performance undoubtedly better. Hence, only a small amount of correlation would be observed between reading scores from the first and the third grade. By the time they finish third grade, most normal readers have attained a good level in reading words, hence a high correlation with their reading scores three years later should be observed. Further, as expected, the most distant measurements (i.e., between kindergarten and sixth grade) are the least correlated.

Finally, as for the phonological measures, we conducted a principal component analysis with varimax rotation on all reading measures (i.e., spelling, word and non-word reading) for each grade. Again, we use the eigenvalue rule of thumb (i.e., eigenvalue >1) as an extraction criteria. In all grades, the three reading scores load between .71 and .95 on one component.

Investigating developmental growth

In this section, we will estimate growth curve models of the different phonological and literacy skills between 1st and 6th grade[1] with the default estimations methods PROC MIXED (for the linear mixed model) in the SAS 9.3 statistical package. These analyses help exploring change over the different testing waves and take into account the non-independence (correlation) of the residuals within a subject (i.e., measures of the same children are correlated). In order to conduct these analyses, we created a person-period data set in which each individual had one record for every testing occasion. In a first step, we will explore the growth shape of each phonological and literacy measure separately using a linear model as reference before investigating the non-linear effects. Next, we will investigate the contribution of the different phonological abilities to reading achievement across the different testing periods.

Results on phonological skills

The same fitting procedure will be followed for each phonological measure. First, we start by fitting an unconditional model. As we do not assume the random-effects covariance matrix to be of any specific form, we specify it to be unstructured. For the residual covariance matrix we use the default 'simple' structure. This means that we assume the residual within-subject variation to be constant and that the inclusion of random effects is sufficient to deal with the non-independence of the residuals within a subject.

Before modeling the covariance structure, we plot the average evolution on each phonological measure for the three different groups (see figures 1.a ; 1.b). The information provided from the exploratory data analysis leads us to add random intercepts and random

---

[1] As the tasks in Kindergarten differed from those used in later grades, scores at kindergarten were not included in the longitudinal analysis.

slopes sequentially for the linear time effect and then for the quadratic one for each measure. Furthermore, we add an interaction between the grouping variable and the significant time effect. In a first step, our primary interest is to compare the average evolution of the phonological scores in the three groups through results on fixed effects. Next, we will test whether the slopes for linear and quadratic time effects are significantly different between groups.

Phonological awareness. The random quadratic effect for both phonological tasks ran into estimation problems hence only the fix effect is introduced in the model. For 'spoonerism', the quadratic effect of time however, is only significant in the groups of normal readers which can be seen in figure 1.a, with a bigger effect for HR-NR group. Furthermore, results show that the linear ($p = .06$) and quadratic ($p = .10$) slopes are not significantly different between groups. In other words, spoonerism does not evolve differently over time in the three groups.

For the 'phoneme deletion' task, a linear and a quadratic time effect are found in normal readers ($p < .05$). However, no significant time effect is found in the dyslexic reader group. In the contrary to spoonerism, both the linear and quadratic slopes are significantly different ($p < .01$) indicating a different evolution across groups. Figure 1.a illustrates well the different growth shapes in each group. It is nonetheless noteworthy that both normal readers share a similarly steep evolution between first and third grade while dyslexic reader' slope is significantly smaller. Conversely, while the low-risk normal readers stop evolving after third grade (due to having reached a ceiling effect) , the high risk group continues to evolve until equalizing the low-risk group's score. Finally, dyslexic readers experience their steepest evolution between third and sixth grade leading to a smaller, yet still significant, difference with the normal readers.

Rapid automatized naming. The graphs displayed in figure 1.b show very similar evolutionary profile for all RAN measures. Hence, we create a composite score with the four RAN measures on which we will fit the growth models. As with PA measures, the random quadratic effect ran into estimation problems hence only the fix effect is introduced in the model. Results show a significant linear ($p < .05$) and quadratic ($p < .001$) trend in all groups. No difference is however significant for the linear ($p = .6$) nor the quadratic

($p = .54$) slope. Hence, as can be seen in figure 1.b, the three groups have the same evolutionary profile with a steeper slope between third and sixth grade than between the first two waves. As no slope difference is observed, the gap between normal and dyslexic readers remains the same over time.

Verbal short-term memory. Results on NRT showed a significant linear effect ($p < .05$) in DR and HR-NR. A quadratic significant effect is also present in all three groups (p's < .01). However, no significant difference in the slopes is found indicating a similar evolution in the three groups. However, in figure one, we can see that both dyslexic and at risk normal readers' slope is slightly higher than the low risk groups, leading to no significant differences between any of the three groups by sixth grade.

Finally, no time effect is found on any of the three groups for the DS task. In addition to the lack of significant difference between the groups on DS, the three groups do not present significantly different slopes reflecting a similar evolution across groups.

Literacy measures. Concerning the 'word-reading test', while normal readers show a linear evolution, dyslexic readers evolve in a quadratic fashion. Additionally, significant differences exist for both linear and quadratic slopes indicating a different evolutionary profile across groups (see figure 2). More precisely, although both groups of normal readers witness a linear evolution, we can see in figure 2 that the high-risk group has a slightly smaller slope between third and sixth grade resulting in a significantly bigger gap between both groups in 6th grade. On the other hand, dyslexic readers evolve in a quadratic fashion with a steeper slope between the second and third wave. The larger slope between third and sixth grade does however not help compensating for the slow evolution between first and third grade which leads to a larger difference between normal and dyslexic readers in sixth grade.

For the 'non-word reading test', the quadratic effect of time is significant in all three groups (p's < .05). This can be verified in figure 2 where the three groups show a steeper evolution between third and sixth grade than between first and third grade. Results also show significant differences in the slopes across the three groups. Figure 2 indeed illustrates that the normal reader low-risk group has a steeper evolution between first and third grade as compared to the two other groups. Consequently, low-risk readers score significantly higher than both other groups in third grade. Conversely, the high-risk group has the steepest evolution of the three groups between third and sixth grade reducing the gap with the low-risk normal readers and maintaining significant difference with the dyslexic readers.

Contribution of phonological ability to reading growth

In this last part, we will investigate whether the three phonological factors (PA, RAN and VSTM) are linked to the variation in the intercept and growth rate of the reading factor score. In that aim, the three phonological scores will be successively added to an unconditional model with the factor score for reading as a dependent variable (Singer & Willet, 2003). We plot the predicted average evolution of the reading factor over time for all participants (see figure 3). The growth line displayed in the graph seems to be slightly curved which suggest that a quadratic growth model might be the best fit for the data at hand. In order to compare models, we first fit an unconditional growth model with a linear time effect, a random intercept and random slope for the linear time effect. As in the previous section, we do not assume the random-effects covariance matrix to be of any specific form so we specify it to be unstructured. We also use the 'simple' structure for the residual covariance matrix.

We then add a quadratic time effect with a random intercept and slope for the quadratic effect. As we run again into estimation problems, we remove the random effect of the quadratic time covariate. The fixed effect, deviance statistic and pseudo R^2 are displayed in the table 5. The Akaike's (AIC) and Schwarz's Bayesian (BIC) information criterion are used to compare the models' goodness of fit. They both penalize for the number of parameters, with BIC giving higher penalty for increased complexity (Singer, 1998). The drop in the AIC and BIC values is very slow and the variance component for growth rate did not decrease indicating that the covariate does not help explaining variation in growth rate.

Correlations conducted earlier have showed the presence of a significant correlation between word-reading and vocabulary in 6th grade. It is then necessary, before investigating the effect of phonological skills, to explore whether the score on vocabulary through the different testing waves have an influence on the overall reading score. We center the vocabulary score around its mean to help interpreting the intercept as the reading score for a person with an average vocabulary score. Results show that the information criterion (AIC and BIC) increase slightly, indicating a worse fit to the data. Furthermore, results on the variance component and growth fit as compared to the unconditional linear growth indicate that the score on vocabulary has no impact on the evolution of the reading score over time. Hence, the unconditional linear growth model will be used as a starting point for the further analyses. Now that we fitted an unconditional growth model, we will explore whether variation in intercepts and slopes are related to the phonological covariates.

We start by incorporating the factor score for PA with an interaction between PA and the time effect. It is noteworthy that, when introducing the grouping variable into the model, we ran into estimation problems most due to the small size of the three groups. Consequently, we will compare the different models based on the average scores on the different factors hence not taking the grouping variable into consideration. Examining the fixed effects shows that the estimates for the intercept and for time are the same, with the difference that they represent now the average intercept and slope 'while controlling for the covariate'. The decrease in the AIC and BIC scores indicates that the new model probably represents a better fit to the data. The coefficient for the covariate (0.10) represents the relation between the covariate and the initial status. As the estimate is almost 4 times bigger than the standard error, we conclude for a significant relation between them. The 'covariate by time' estimate indicates that individuals who differ by 1 with respect to PA have a grow rate that differs by 0.03. However this is not significant hence, based on the data at hand, we cannot prove that reading skills evolve differently in function of PA. However, these results have to be carefully interpreted since, as we saw in the previous section, differences in growth were observed for one of PA tasks (i.e., 'phoneme deletion') but not for the other (i.e.,'spoonerism'). The residual estimate has remained unchanged. We do nonetheless observe a small decrease in the fit of growth (i.e., the estimates for the variance-covariance matrix for the slopes) as compared to the unconditional model (from 0.03 to 0.02). Computing (0.03-0.02)/0.03=0.33 indicates that the covariate accounts for 33% of the explained variation in growth rate. Hence, including PA did improve the fit of the growth rates.

The next covariate that we incorporate to the unconditional model is RAN. Here again, comparing goodness of fit indicates that the model that includes the covariate better fits the data than the unconditional model. The Coefficient for the covariate (0.05) indicates no significant relation with the initial status. Further, the interaction between the covariate and time indicates that individuals who differ by one with respect to RAN have a growth rate in reading that significantly differs by 0.10 in function of RAN's score. However, the interaction being non-significant, we cannot say that reading skills evolve differently across children who score differently at RAN which corresponds to the absence of differences in the evolution of RAN observed in the previous section. The variance component for the intercept slightly increases signifying that the variation in the intercept and slope cannot be explained by the covariate RAN. However, including RAN did improve the fit of growth rate: compared to the initial model, RAN does account for a significant amount (42%) of the variance in growth rate. The non-significant estimate for the residual indicates that there is not much variability in reading that is left to be explained.

Finally, adding VSTM does not provide a better fit to our data. Furthermore, the intercept for the covariate does not show a significant relation with the initial status. Surprisingly we do find a small, yet significant, effect of the covariate with respect to growth rate ($p < .01$). Hence, the individuals who differ by 1 with respect to VSTM, have a growth rate that differs by 0.03. Finally, VSTM does not account for the growth rate's variance.

Discussion

In this study, children were tested on phonological and literacy skills from kindergarten to 3rd grade. The same participants were tested again when they have reached 6th grade with the aim of further exploring the evolution of these skills. The addition of a testing wave has lead us to modify the criteria for including participants in a 'dyslexic' group. Hence, a child was categorized as a dyslexic reader if he/she had scored lower than percentile 10 on at least two measurement moments for the spelling and/or reading tasks and lower than percentile 50 for both reading and spelling tasks on all time points. As expected, taking into consideration 6th grade's literacy scores to categorize dyslexic readers lead to sharper differences between groups than what was found in our earlier analysis. More precisely, significant differences between the DR and normal readers for PA measures (i.e., ‘rhyme production’ and ‘first-sound’ identity’) were found in kindergarten indicating that a phonological deficit might exist within dyslexic readers before the first onset of reading.

The first aim of our study was to examine whether dyslexia is an all-or-none condition or rather a continuum where the severity of the cognitive and literacy difficulties would vary between individuals. Results are in congruence with previous findings where normal readers at high risk of dyslexia scored better than dyslexic readers but yet lower than normal readers at low risk which confirms the presence of a continuum in the readers’ cognitive profiles where the phonological deficit can have different levels of difficulty that would vary in function of co-occuring risk factors (Snowling, Gallagher & Frith, 2003; Peterson, Pennington, Shirberg & Boada, 2009). This theory is supported by other studies showing that while some dyslexic readers do not show a phonological deficit, other readers with phonological difficulties are normal readers (Moll, Loff & Snowling, 2013).

In agreement with existing literature (e.g., Menghini, Finzi, Benassi, Bolzani, Facoetti, Giovagnoli, Ruffinod, Vicari, 2010; Ramus et al., 2003; Snowling, 2000), a phonological deficit was observed in most phonological domains at all time points. More precisely, group comparison showed that dyslexic readers scored lower than normal readers on RAN and PA at all time-points. Regarding VSTM, in agreement with previous reports; significant differences between dyslexic and normal readers were found on the non-word repetition task between kindergarten and third grade (e.g., Moll, Loff & Snowling, 2013; van Bergen et al., 2012). No significant differences were however found in sixth grade. This lack of deficit on the verbal short-term memory in 6th grade measure can be due to a compensation mechanism developed by dyslexic readers to balance for the phonological difficulties. The ‘digit span’ task however, did not help discriminating between dyslexic and normal readers at no time point. Although unexpected, these results are in congruence with previous reports on dyslexic readers where memory capacities seemed to be preserved (Menghini, Finzi, Benassi et al., 2010; Parrila, Kirby & McQuarrie, 2004). While these convergent results may be due to the heterogeneity of tasks and sample selection, it can be argued that they reflect the complexity of developmental dyslexia which leads to heterogeneity of profiles implicating different cognitive deficits (Menghini, Finzi, Benassi et al., 2010; Pennington, 2006).

Finally, results on the different phonological measures confirm the presence of a general phonological deficit that is present before the formal onset of reading and that is persistent even after five years of reading instructions. These findings are in line with previous longitudinal studies that compared the evolution of children at high and low risk of developing dyslexia (Pennington & Lefly, 2001; Snowling et al., 2003). Moreover, the fact that unaffected at-risk children scored lower than those from low risk families on the phonological tasks (although these differences did not appear to be statistically significant), goes in line with previous findings suggesting that the phonological deficit is an endophenotype of dyslexia (van Bergen, de Jong, Plakas, Maassen, & van der Leij, 2012; Boets et al., 2010; Snowling, Muter & Carroll, 2007).

In a second step, we compared the different groups on their literacy skills. Dyslexic readers scored significantly lower than both normal reader groups for each literacy task at all time points. As for phonological processes, the high-risk normal readers scored lower than the low-risk group at all literacy tasks and for each time point confirming that the family risk for dyslexia is continuous rather than discrete. These results are in accordance with other longitudinal studies of children at family risk of dyslexia (e.g., Pennington & Lefly, 2001; Snowling et al., 2003). Importantly, we can observe a switch in the 'literacy profile' of the high-risk normal readers group: in third grade, the high-risk normal readers scored significantly lower than the low-risk normal readers on both non-word and spelling tasks. These differences however –although still existent- become statically non-significant by 6th grade. Furthermore, at this stage, an unexpected significant difference appears between the two groups for the real word reading task. These results can be explained by the evolutionary profiles of the two normal reader's groups. More precisely, the high-risk group presents a quadratic growth as compared to the low-risk which illustrates a linear one. Nevertheless, while the slope for non-word reading task becomes much steeper between third and sixth grade reducing the gap with the low risk group, it decreases a little bit for the real word reading task leading to a more significant gap between both groups.

Next, we aimed at comparing growth trajectories of both phonological and literacy skills between the three reading groups. Interestingly, results showed different developmental profiles for literacy tasks as well as for phoneme deletion. On the other hand, no differences in the slopes for the other tasks were found. It is noteworthy that the tasks where the three groups exposed similar developmental profiles all required - to a certain degree - the solicitation of memory functions. More precisely, similar slopes were observed for both VSTM tasks, but also for RAN which implies the retrieval of phonological codes for long term memory and 'spoonerism' which is a phonological task that demands to maintain two words (or non-words) while accomplishing the task hence soliciting working memory. Consequently, the gap between the three groups stays more or less constant between first and 6th grade, except for literacy scores where the gap seems to widen by 6th grade.

In line with what has been observed in previous studies, the three groups mostly followed a quadratic developmental trajectory. These results are in accordance with the 'deficit-model', where good readers have been observed to have a higher intercept than dyslexic readers while they follow the same quadratic trajectory (Wei, Blackorby, & Schiller, 2011; Catts, Bridges, Little, & Tomblin, 2008). Hence, as it has been observed in previous reports (e.g., Bruck, 1992; Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996), dyslexic readers fail to develop adequate reading even after 6years of reading instruction.

Finally, our last aim was to explore the relation between phonological and literacy skills. Consistent with other studies (e.g., Verhagen, Aarnoutse, & Van Leeuwe, 2010), PA and RAN explained a significant amount of variance in the evolution of reading with PA having a bigger effect. After including PA and RAN, the estimate for the residual becomes non-significant indicating that there is not much variability in reading left to be explained. Hence, when we introduce VSTM to the model, it does not account for the growth rate's variance. While PA significantly correlated with the initial status on reading, no such effect was found for RAN nor VSTM. Surprisingly we do find a small, yet significant, effect of the covariate with respect to growth rate. Hence, the individuals who differ by 1 with respect to VSTM, have a grow rate that differs by 0.03.

Results on this study were in accordance with previous studies (e.g., de Jong and van der Leij, 2002; Landerl and Wimmer ,2008), PA and RAN constitute powerful predictors to reading ability in 6th grade. It is however noteworthy that the grouping variable could not be introduced in the model due to the small sample size of the different groups. Hence, a similar study with a larger sample size would be needed to further compare the prediction power of each factor.

Conclusion

This study confirmed the presence of a persistent phonological deficit within children with dyslexia and this before the first onset of reading and after 6 years of formal reading instructions. Although this deficit was present in all three factors until third grade, no deficit is observed in VSTM in 6th grade probably reflecting the development of a compensating mechanism. Further, children with dyslexia showed impairment in all reading measures through all time points. Additionally, high-risk normal readers score higher than dyslexic readers on both phonological and reading tasks yet lower than low-risk normal readers suggesting that the family risk of dyslexia is continuous rather than discrete. Finally, only PA and RAN appeared to significantly predict reading performance through 6th grade. Finally, although VSTM did not significantly explain variation of growth rate in reading, its significant interaction with time does suggest that reading skills evolve differently across children with different VSTM abilities.

Reference List

Aarnoutse, C., van Leeuwe, J., Voeten, M., & Oud, H. (2001). Development of decoding, reading comprehension, vocabulary and spelling during the elementary school years. *Reading and writing, 14,* 61-89.

Blomert, L., & Willems, G. (2010). Is there a causal link from a phonological awareness deficit to reading failure in children at familial risk for dyslexia? *Dyslexia, 16,* 300-317.Boets, B., De, S. B., Cleuren, L., Vandewalle, E., Wouters, J., & Ghesquiere, P. (2010). Towards a further characterization of phonological and literacy problems in Dutch-speaking children with dyslexia. *British Journal of Developmental Psychology, 28,* 5-31.

Boets, B., Vandermosten, M., Poelmans, H., Luts, H., Wouters, J., & Ghesquiere, P. (2011). Preschool impairments in auditory processing and speech perception uniquely predict future reading problems. *Research In Developmental Disabilities, 32,* 560-570.

Bowers, P. G. (1995). Tracing symbol naming speed's unique contributions to reading disabilities over time. *Reading and Writing,* 7(2), 189–216.

Brus, B. T., & Voeten, M. J. (1973). *Eén-minuut-test vorm A en B. Verantwoording en handleiding* [*One-minute word reading test version A and B. Justification and manual*]. Nijmegen, the Netherlands:Berkhout.

Catts, H. W., Bridges, M. S., Little, T. D., & Tomblin, J. B. (2008). Reading Achievement Growth in Children With Language Impairments. *Journal of speech language and hearing research, 51*, 1569-1579.

Catts, H. W., Gillispie, M., Leonard, L. B., Kail, R. V., & Miller, C. A. (2002). The role of speed of processing, rapid naming, and phonological awareness in reading achievement. *Journal of Learning Disabilities, 35,* 509-524.

Caravolas, M., Hulme, C., & Snowling, M. J. (2001). The foundations of spelling ability: Evidence from a 3-year longitudinal study. *Journal of Memory and Language, 45*, 751‐774. doi:10.1006/jmla.2000.2785.

Caravolas, M., Lervag, A., Mousikou, P., Efrim, C., Litavsky, M., Onochie-Quintanilla, E., Salas, N., Schöffelová, M., Defior,S., Mikulajová, M., Seidlová-Málková, G., & Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. Psychological Science, 23, 678–686.

Chiappe, P., Hasher, L., & Siegel, L. S. (2000). Working memory, inhibitory control, and reading disability. *Memory and Cognition, 28,* 8-17.

Denckla, M. B. & Cutting, L. E. (1999). History and significance of rapid automatized naming. *Annals of Dyslexia, 49,* 29-42.

Denckla, M. B. & Rudel, R. G. (1976). Rapid "automatized" naming (R.A.N): dyslexia differentiated from other learning disabilities. *Neuropsychologia, 14,* 471-479.

Dudal,P.(1997). Leerlingvolgsysteem VCLB (CSBO), Spelling: Toetsen1-2-3. Basisboekenkopieerbundel [Studenttrajectorysystem.SpellingGrade1-2-3.Manual] . Garant: Leuven.

de Bree, E., Wijnen, F., & Gerrits, E. (2010). Non-word repetition in Dutch children at-risk of dyslexia and children with SLI: results of the follow-up study. *Dyslexia, 16,* 36-44.

de Jong, P. F., & van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology, 91,* 450-476.

de Jong, P. F., Seveke, M. J., & van Veen, M. (2000). Phonological sensitivity and the acquisition of new words in children. Journal of Experimental Child Psychology, *76*, 275–301.

de Jong, P. F., & van der Leij, A. (2002). Effects of phonological abilities and linguistic comprehension on the development of reading. *Scientific Studies of Reading*, *6*, 51–77.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology, 88,* 3-17.

Francks, C., MacPhie, I. L., & Monaco, A. P. (2002). The genetic basis of dyslexia. *Lancet Neurology, 1,* 483-490.

Froyen, D., Willems, G., & Blomert, L. (2011). Evidence for a specific cross-modal association deficit in dyslexia: an electrophysiological study of letter-speech sound processing. *Developmental Science, 14,* 635-648.

Georgiou, G.K., Parrila, R., & Papadopoulos, T.C. (2008). Predictors of word decoding and reading fluency across languages varying in orthographic consistency. Journal of Educational Psychology, 100, 566–580.

Gersons-Wolfensberger, D. C. & Ruijssenaars, W. A. (1997). Definition and treatment of dyslexia: a report by the Committee on Dyslexia of the Health Council of The Netherlands. *Journal of Learning Disabilities, 30*, 209-213.

Gilger, J., Pennington, B., & DeFries, J. C. (1991). Risk for Reading Disability as a Function of Parental History in Three Family Studies. In B.Pennington (Ed.), *Reading Disabilities* (4 ed., pp. 17-29). Springer Netherlands.

Griffiths, Y. M. & Snowling, M. J. (2002). Predictors of Exception Word and Nonword Reading in Dyslexic Children: The Severity Hypothesis. *Journal of Educational Psychology, 94*, 34-43.

Howell, D., C. (2010). Statistical methods for psychology (7th ed.). Belmont, CA : Thomson Wadsworth.

Jeffries, S. & Everatt, J. (2004). Working memory: its role in dyslexia and other specific learning difficulties. *Dyslexia., 10,* 196-214.

Jordan, N. C., Kaplan, D., & Hanich, L. B. (2002). Achievement Growth in Children With Learning Difficulties in Mathematics. *Journal of Educational Psychology, 94,* 586-597.

Katz, R. B., Shankweiler, D., & Liberman, I. Y. (1981). Memory for item order and phonetic recording in the beginning reader. *Journal of Experimental Child Psychology, 32,* 474-484.

Kibby, M. Y. (2009). There are multiple contributors to the verbal short-term memory deficit in children with developmental reading disabilities. *Child Neuropsychology, 15,* 485-506.

Kibby, M. Y. & Cohen, M. J. (2008). Memory functioning in children with reading disabilities and/or attention deficit/hyperactivity disorder: a clinical investigation of their working memory and long-term memory functioning. *Child Neuropsychology., 14,* 525-546.

Kibby, M. Y., Marks, W., Morgan, S., & Long, C. J. (2004). Specific impairment in developmental reading disabilities: a working memory approach. *Journal of Learning Disabilities, 37,* 349-363.

Kirby, J. R., Parrila, R., & Pfeiffer, S. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology, 95,* 453–464.

Kipp, K. H. & Mohr, G. (2008). Remediation of developmental dyslexia: tackling a basic memory deficit. *Cognitive Neuropsychology., 25,* 38-55.

Kirby, J. R., Silvestri, R., Allingham, B. H., Parrila, R., & La Fave, C. B. (2008). Learning strategies and study approaches of postsecondary students with dyslexia. *Journal of Learning Disabilities., 41,* 85-96.

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

Landerl, K., Ramus, F., Moll, K. Lyytinen, L., e. a. (2013). Predictors of developmental dyslexia in European orthographies with varying complexity. *Journal of Child Psychology and Psychiatry, 54,* 686–694.Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A Definition of Dyslexia. *Annals of Dyslexia, 53,* 1-14.

Mann, V. A. & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory. *Journal of Learning Disabilities, 17,* 592-599.

Maughan, B. (1995). Annotation: long-term outcomes of developmental reading problems. *Journal of Child Psychology and Psychiatry, 36,* 357-371.

Menghini, D., Finzi, A., Benassi, M., Bolzani, R., Facoetti, A., Giovagnoli, S. et al. (2010). Different underlying neurocognitive deficits in developmental dyslexia: A comparative study. *Neuropsychologia, 48,* 863-872.

Moll, K., Loff, A., & Snowling, M. (2013). Cognitive endophenotypes of dyslexia. *Scientific Studies of Reading, 00,* 1–13.

Nakamoto, J., Lindsey, K. A., & Manis, F. R. (2006). A Longitudinal Analysis of English Language Learners' Word Decoding and Reading Comprehension. *Reading and Writing: An Interdisciplinary Journal, 20,* 691-719.

Parilla, R., Kirby, J.R., & McQuarrie, L. (2004). Articulation rate, naming speed, verbal short-term memory, and phonemic awareness: Longitudinal predictors of early reading development? Scientific Studies of Reading, 8, 3‑26.

Peterson, R. L., Pennington, B. F., Shriberg, L. D., & Boada, R. (2009). What influences literacy outcome in children with speech sound disorder? *Journal of Speech, Language, and Hearing Research, 52,* 1175–1188.

Peterson, R. L. & Pennington, B. F. (2012). Developmental dyslexia. *The Lancet, 379,* 1997-2007.

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition, 101,* 385–413.

Pennington, B. F. & Lefly, D. L. (2001). Early Reading Development in Children at Family Risk for Dyslexia. *Child Development, 72,* 816-833.

Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Leppänen, H. T., Poikkeus, A., Tolvanen, A., Torppa, M., & Lyytinen, H. (2008). Developmental links of very early phonological and language skills in second grade: reading outcomes. *Journal of Learning Disabilities, 41,* 353-370.Ramus, F. & Szenkovits, G. (2008). What phonological deficit? *Quarterly Journal of Experimental Psychology.(Hove.), 61,* 129-141.

Savage, R., Pillay, V., & Melidona, S. (2008). Rapid serial naming is a unique predictor of spelling in children. *Journal of Learning Disabilities, 41,* 235-250.

Savage, R. S. & Frederickson, N. (2006). Beyond phonology: What else is needed to describe the problems of below-average readers and spellers? *Journal of Learning Disabilities, 39*, 399-413.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology, 96,* 265-282.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of educational and behavioral statistics, 23*, 323-355.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology, 96,* 265-282.

Shaywitz, S. E. & Shaywitz, B. A. (2003). The science of reading and dyslexia. *Journal of AAPOS, 7,* 158-166.

Snowling, M. (2000). Dyslexia (2nd ed.). Oxford: Blackwell.

Snowling, M., Bishop, D. V., & Stothard, S. E. (2000). Is preschool language impairment a risk factor for dyslexia in adolescence? *Journal of Child Psychology and Psychiatry, 41,* 587-600.

Snowling, M. J., Gallagher, A., & Frith, U. (2003). Family risk of dyslexia is continuous: individual differences in the precursors of reading skill. *Child Development, 74,* 358-373.

Snowling, M.J., Muter, V., & Carroll, J. (2007). Children at family risk of dyslexia: A follow-up in early adolescence. *Journal of Child Psychology & Psychiatry*, *48,* 609-618.

Stanovich, K. E., Nathan, R. G., & Zolman, J. E. (1988). The Developmental Lag Hypothesis in Reading: Longitudinal and Matched Reading-Level Comparisons. *Child Development, 59*, 71-86.

Swanson, H. L., Kehler, P., & Jerman, O. (2010). Working memory, strategy knowledge, and strategy instruction in children with reading disabilities. *Journal of Learning Disabilities, 43*, 24-47.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997). Contributions of Phonological Awareness and Rapid Automatic Naming Ability to the Growth of Word-Reading Skills in Second-to Fifth-Grade Children. *Scientific Studies of Reading, 1,* 161-185.

van den Bos, K.P., Spelberg, H.C.L., Scheepstra, A.J.M., &DeVries, J.R. (1994). De Klepel.VormAenB.Eentestvoordeleesvaardigheidvanpseudowoorden.Verantwoording, handleiding,diagnostiekenbehandeling. Berkhout:Nijmegen.

van den Bos, K. P., Zijlstra, B. J. H., & Lutje Spelberg, H. C. (2002). Life-Span Data on Continuous-Naming Speeds of Numbers, Letters, Colors, and Pictured Objects, and Word-Reading Speed. *Scientific Studies of Reading, 6,* 25-49.

Vandewalle, E., Boets, B., Ghesquière, P., & Zink, I. (2010). Who is at risk for dyslexia? Phonological processing in five- to seven-year-old Dutch-speaking children with SLI. *Scientific Studies of Reading, 14,* 58-84.

van Bergen, E., de Jong, P. F., Plakas, A., Maassen, B., & van der Leij, A. (2012). Child and parental literacy levels within families with a history of dyslexia. *Journal of Child Psychology & Psychiatry, 53*, 28-36.

van Bergen, E. (2013). *Who will develop dyslexia? Cognitive precursors in parents and children*. Dissertatie. Universiteit van Amsterdam.

Vaessen, A., Bertrand, D., To´th, D., Csepe, V., Faı´sca, L., Reis, A., & Blomert, L. (2010). Cognitive development of fluent word reading does not qualitatively differ between transparent and opaque orthographies. Journal of Educational Psychology, 102, 827–842.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,* 2-40.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R. S. et al. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology* 88[4], 601-638.

Verhagen, W. G. M., Aarnoutse, C. A. J., & Van Leeuwe, J. F. J. (2010). Spelling and word recognition in Grades 1 and 2: Relations to phonological awareness and naming speed in Dutch children. *Applied Psycholinguistics, 31,* 59-80.

Wagner, R. K. & Torgesen, J. K. (1987). The Nature of Phonological Processing and Its Causal Role in the Acquisition of Reading Skills. *Psychological Bulletin, 101,* 192-212.Wagner, R.K., Torgesen, J.K., & Rashotte, C.A. (1999). Comprehensive Test of Phonological Processing (CTOPP). Austin, TX: PRO-ED.

Wei, X., Blackorby, J., & Schiller, E. (2011). Growth in Reading Achievement of Students with Disabilities, Ages 7 to 17. *Exceptional Children, 78,* 89-106.

Wimmer, H., Mayringer, H., & Landerl, K. (2000). The doubledeficit hypothesis and difficulties in learning to read a regular orthography. *Journal of Educational Psychology, 92*, 668‐680.

Wimmer, H. & Schurz, M. (2010). Dyslexia in regular orthographies: manifestation and causation. *Dyslexia., 16,* 283-299.

Wolf, M., Bally, H., & Morris, R. (1986). Automaticity, retrieval processes, and reading: a longitudinal study in average and impaired readers. *Child Development, 57,* 988-1000.

Wolf, M. & Bowers, P. G. (2000). Naming-speed processes and developmental reading disabilities: an introduction to the special issue on the double-deficit hypothesis. *Journal of Learning Disabilities, 33*, 322-324.

Wolf, M., Bowers, P. G., & Biddle, K. (2000). Naming-Speed Processes, Timing, and Reading: A Conceptual Review. *Journal of Learning Disabilities, 33,* 387-407.

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

Table 1: Cronbach alpha and principal component analysis for 6th grade with varimax rotation: factor loadings of the phonological measures.

	Cronbach's alpha	Factor I, PA	Factor 2, RAN	Factor 3, VSTM
Kindergarten measures				
Rhyme production	(.88)	.68		
Rhyme identity	(.69)	.83		
Firs-sound identity	(.59)	.82		
End-sound identity	(.63)	.80		
Colour naming			.91	
Picture naming			.92	
Digit span				.83
Nonword repetition				.74
First grade measures				
Rhyme categorization	(.69)	.83		
First-sound categorization	(.71)	.74		
Phoneme deletion	(.84)	.74		
Spoonerism	(.91)	.80		
Colour naming			.83	
Picture naming			.84	
Digit naming			.90	
Letter naming			.81	
Digit span				.82
Nonword repetition	(.84)			.84
Third grade measures				
Phoneme deletion	(.92)	.88		
Spoonerism	(.92)	.89		
Colour naming			.84	
Picture naming			.75	
Digit naming			.86	
Letter naming			.84	
Digit span				.90
Nonword repetition	(.79)			.84
Sixth grade measures				
Phoneme deletion	(.79)	.87		
Spoonerism	(.90)	.89		
Colour naming			.82	
Picture naming			.80	
Digit naming			.86	
Letter naming			.75	
Digit span				.87
Nonword repetition	(.78)			.88

Note: Solely the factor loadings >.65 were accepted and reported in the table

Table 2. a: Descriptive statistics on phonological measures for the three participant groups at kindergarten.

	Maximum	DR		HR-NR		LR-NR	
		M	SD	M	SD	M	SD
Kindegarten measures							
Factor PA		-0.74$_{a}$	1.20	-0.63$_{ab}$	1.19	0.00$_{b}$	1.00
Rhyme production	12	5,92$_{a}$	2,66	6,57$_{ab}$	2,23	7,21$_{b}$	1,85
Rhyme identity	12	8,31$_{a}$	3,011	8,95$_{ab}$	2,42	9,96$_{b}$	1,81
First-sound identity	10	4,23$_{a}$	2,13	4,90$_{b}$	2,23	5,64$_{b}$	2,30
End-sound identity	10	4,31$_{a}$	1,80	4,62$_{ab}$	2,48	6,07$_{b}$	2,37
Factor RAN		-1$_{a}$	1.02	-0.27$_{ab}$	1.09	0.00$_{b}$	1.00
Color naming	-	0,56$_{a}$	0,07	0,65$_{ab}$	0,13	0,71$_{b}$	0,16
Object naming	-	0,57$_{a}$	0,10	0,65$_{ab}$	0,13	0,71$_{b}$	0,16
Factor VSTM		-0.30	0.74	-0.19	0.93	0.00	1.00
Digit span	10	6,69	1,60	7,33	1,59	6,82	1,49
Nonword repetition	32	16,23$_{a}$	5,69	16,90$_{a}$	5,87	20,89$_{b}$	6,65

Note: Pairs with different subscript letters differ significantly (univariate MMA, Planned contrasts, $p<.001$- one-tailed).

Table 2. b: Descriptive statistics on phonological measures for the three participant groups at 1st grade.

	Maximum	DR		HR-NR		LR-NR	
		M	SD	M	SD	M	SD
Factor PA		-0.95$_{a}$	0.83	-0.12$_{b}$	1.12	0.00$_{b}$	1.00
Phoneme deletion	19	6,3$_{a}$	3,42	11,05$_{b}$	4,61	11,64$_{b}$	4,64
Spoonerism	43	13,92$_{a}$	7,30	18,24$_{ab}$	10,88	21,50$_{b}$	8,60
Factor RAN		-1.64$_{a}$	0.76	-0.18$_{b}$	0.97	0.00$_{b}$	1.00
Color naming	-	0,70$_{a}$	0,15	0,89$_{b}$	0,18	0,91$_{b}$	0,18
Object naming	-	0,68$_{a}$	0,11	0,87$_{b}$	0,18	0,89$_{b}$	0,15
Digit naming	-	0,81$_{a}$	0,22	1,19$_{b}$	0,21	1,28$_{b}$	0,27
Letter naming	-	0,90$_{a}$	0,16	1,29$_{b}$	0,4	1,38$_{b}$	0,28
Factor VSTM		-0.36	0.78	-0.31	1	0.00	1.00
Digit span	13	8,31	1,18	8,67	1,77	9,00	1,47
Nonword repetition	35	19,85$_{a}$	4,30	21,14$_{ab}$	5,75	23,21$_{b}$	6,21

Note: Pairs with different subscript letters differ significantly (univariate MMA, Planned contrasts, $p<.001$-one-tailed).

Table 2. c: Descriptive statistics on phonological measures for the three participant groups at 3rd grade.

	Maximum	DR		HR-NR		LR-NR	
		M	SD	M	SD	M	SD
Factor PA		-1.23$_a$	0.80	-0.40$_b$	0.88	0.00$_b$	1.00
Phoneme deletion	28	7,85$_a$	4,36	17,57$_b$	5,59	20,36$_b$	6,61
Spoonerism	49	22,$_a$	10,50	32,29$_b$	9,32	37,21$_b$	8,77
Factor RAN		-1.15$_a$	0.66	-0.19$_b$	0.88	0.00$_b$	1.00
Color naming	-	0,84$_a$	0,14	1,05$_b$	0,21	1,05$_b$	0,20
Object naming	-	0,84$_a$	0,13	0,98$_b$	0,16	0,98$_b$	0,17
Digit naming	-	1,16$_a$	0,26	1,64$_b$	0,28	1,71$_b$	0,28
Letter naming	-	1,15$_a$	0,23	1,46$_b$	0,24	1,51$_b$	0,34
Factor VSTM		-0.43	1.15	-0.26	1.17	0.00	1.00
Digit span	15	9,00	1,68	9,62	1,83	9,75	1,58
Nonword repetition	39	23,15$_a$	5,08	24,14$_{ab}$	5,45	26,79$_b$	5,97

Note: Pairs with different subscript letters differ significantly (univariate MMA, Planned contrasts, $p<.001$-one-tailed).

Table 2. d: Descriptive statistics on phonological measures for the three participant groups at 6thy grade.

	Maximum	DR		HR-NR		LR-NR	
		M	SD	M	SD	M	SD
Factor PA		$-2{,}39_a$	2,03	$-0{,}11_b$	1,04	0.00_b	1.00
Phoneme deletion	24	$13{,}62_a$	5,36	$19{,}16_b$	2,69	$19{,}60_b$	2,65
Spoonerism	48	$24{,}08_a$	11,60	$37{,}50_b$	5,57	$38{,}48_b$	6,82
Factor RAN		$-1{,}20_a$	1,02	$-0{,}17_b$	1,21	0.00_b	1.00
Object naming	-	$1{,}06_a$	0,18	$1{,}22_b$	0,20	$1{,}28_b$	0,17
Digit naming	-	$1{,}88_a$	0,34	$2{,}28_b$	0,48	$2{,}29_b$	0,38
Color naming	-	$1{,}21_a$	,0249	$1{,}40_b$	0,23	$1{,}42_b$	0,23
Letter naming	-	$1{,}82_a$	0,35	$2{,}31_b$	0,43	$2{,}34_b$	0,37
Factor VSTM		-0,37	1,40	-,5158	1,05	0.00	1.00
Digit span	19	10,38	2,40	11,05	2,59	10,97	2,18
Nonword repetition	44	32,38	5,75	32,65	5,09	34,60	4,51

Note: Pairs with different subscript letters differ significantly (univariate MMA, Planned contrasts, p<.001-one-tailed).

Table 3: descriptive statistics on the literacy measures for the three groups.

	Maximum	DR		NR-HR		NR-LR	
		M	SD	M	SD	M	SD
Kindergarten							
Letter knowledge	30	4,31$_a$	3,96	6,62	7,72	8,29$_b$	6,94
First grade measures							
Word reading	51	8,54$_a$	4,35	19,62$_b$	6,22	22,68$_b$	9,10
Non-word reading	40	8,62$_a$	3,59	20,62$_b$	7,52	23,21$_b$	9,02
Spelling (standard score)	-	37,23$_a$	13,86	51,00$_b$	7,08	53,50$_b$	4,61
Third grade measures							
One-minute reading		20,00$_a$	11,90	43,67$_b$	10,96	48,71$_b$	12,91
Non-word reading	70	13,85$_a$	8,21	34,05$_c$	10,72	43,04$_b$	13,33
Spelling (standard score)	-	14,69$_a$	5,60	21,67$_c$	3,76	24,29$_b$	3,47
Sixth grade measures	-						
One-minute reading	101	45,00$_a$	15.85	66,10$_c$	13,39	74,56$_b$	10,65
Non-word reading	92	37,69$_a$	16,69	64,55$_b$	12,27	70,96$_b$	12,45
Spelling (standard score)	-	11,08$_a$	6,50	19,70$_b$	2,77	21,64$_b$	3,33

Note 1: Pairs with different subscript letters differ significantly
Note 2: Standardized scores with a population average M=100 and SD=15. These literacy measures were used to define the dyslexic and normal reading groups.

Table 4: Pearson correlations for literacy measures

	1st grade	3rd grade
Word reading		
3rd grade	.72**	
6th grade	.55**	.78**
Non-word reading		
3rd grade	.72**	
6th grade	.65**	.85**
Spelling		
3rd grade	.50**	
6th grade	.43*	¨.75**

**p<.001

Table 4: Comparing model fit

Reading				
	Unconditional linear growth	**PA**	**RAN**	**VSTM**
AIC	136.4	128.5	126.2	143.5
BIC	144.9	137.0	134.7	152
Linear term	1.09***	1.11***	1.09	1.09***
Intercept for covariate		0.10***	0.05	-0.04
Covariate*time		0.030	0.10	0.03**
UN(1.1) Variance component for the intercept	0.035**	0.01255	0.02	0.04
UN (1.2)	0.039***	0.03769***	0.03	0.04
UN (2.2) Variance for growth rate	0.033**	0.02	0.019	0.03
Residual	0.05***	0.06***	0.05	0.05
Intercept	-1.06***	-1.07***	-1.07	-1.07***

*p<.05;**p<.01;***p<.001

Figure1.a : Predicted average evolution of phonological scores over time in the group of dyslexic readers (DR), normal readers at high risk of dyslexia (HR-NR) and normal readers at low risk of dyslexia (LR-NR).

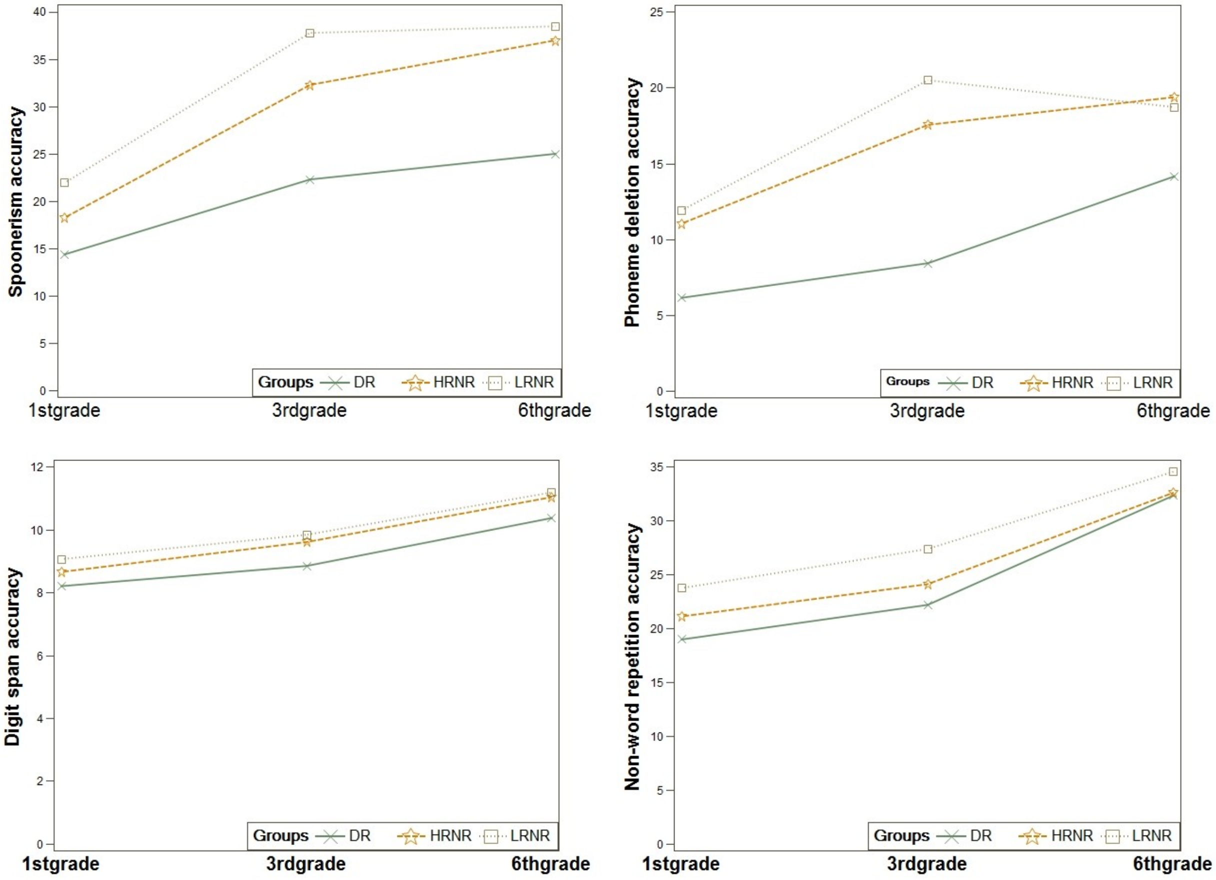

Figure1.b : Predicted average evolution of RAN scores over time in the group of dyslexic readers (DR), normal readers at high risk of dyslexia (HR-NR) and normal readers at low risk of dyslexia (LR-NR).

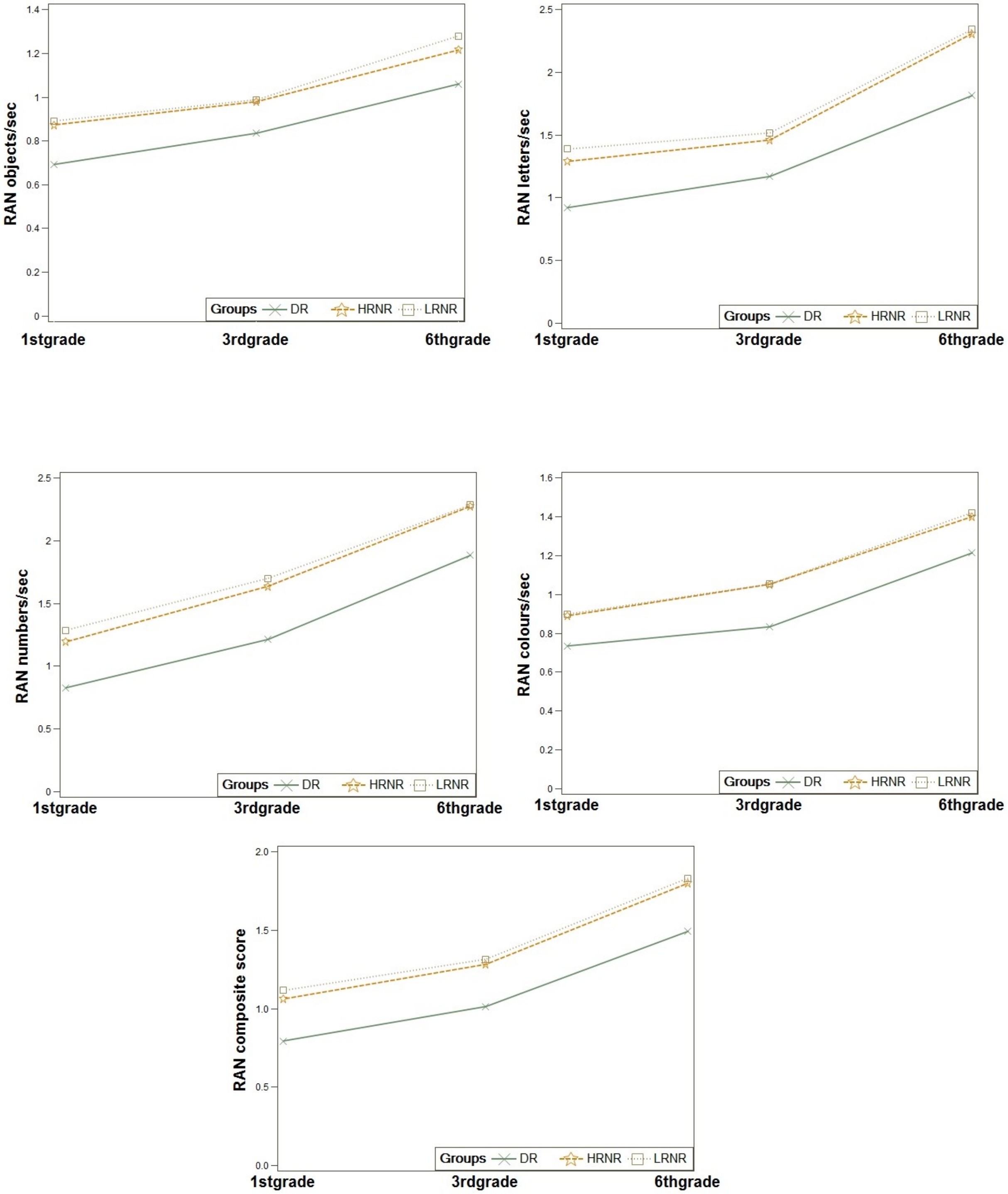

Figure2: Predicted average evolution of the score for word-reading and non-word reading over time in the group of dyslexic readers (DR), normal readers at high risk of dyslexia (NRHR) and normal readers at low risk of dyslexia (NRLR).

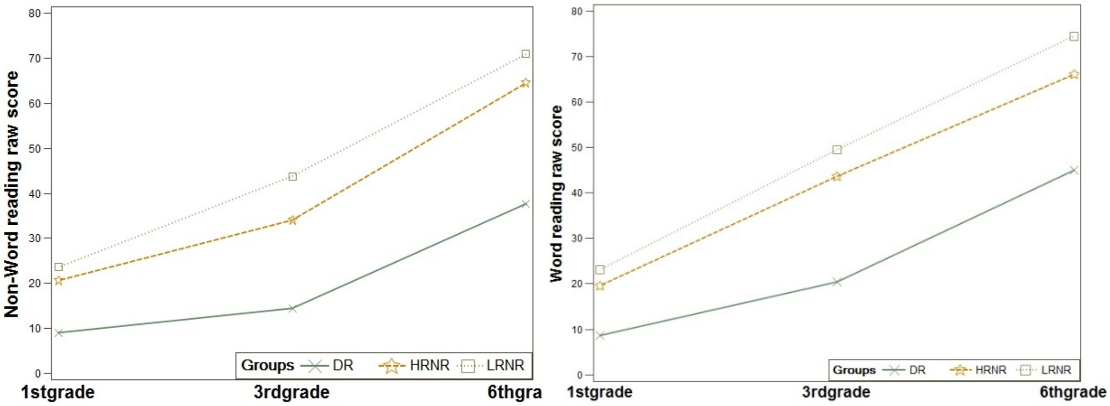

Figure 3: Predicted average evolution of the factor score for reading over time in the group of dyslexic readers (DR), normal readers at high risk of dyslexia (HR-NR) and normal readers at low risk of dyslexia (LR-NR).
Note: grade 1= time 0; grade 3= time1; grade6= time 2.