

**This is an electronic reprint of the original article.
This reprint *may differ* from the original in pagination and typographic detail.**

Author(s): Torppa, Minna; Eklund, Kenneth; van Bergen, Elsje; Lyytinen, Heikki

Title: Late-Emerging and Resolving Dyslexia: A Follow-Up Study from Age 3 to 14

Year: 2015

Version:

Please cite the original version:

Torppa, M., Eklund, K., van Bergen, E., & Lyytinen, H. (2015). Late-Emerging and Resolving Dyslexia: A Follow-Up Study from Age 3 to 14. *Journal of Abnormal Child Psychology*, 43(7), 1389-1401. <https://doi.org/10.1007/s10802-015-0003-1>

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Running head: LATE-EMERGING AND RESOLVING DYSLEXIA

**LATE-EMERGING AND RESOLVING DYSLEXIA:
A FOLLOW-UP STUDY FROM AGE 3 TO 14**

PhD Minna Torppa (University of Jyväskylä, Finland, minna.p.torppa@jyu.fi),
Kenneth Eklund, MA (University of Jyväskylä, Finland, kenneth.m.eklund@jyu.fi),
PhD Elsje van Bergen (University of Oxford, UK, elsje.vanbergen@psy.ox.ac.uk),
and
Prof. Heikki Lyytinen (University of Jyväskylä, Finland, heikki.lyytinen@jyu.fi)

Torppa, M., Eklund, K., van Bergen E., & Lyytinen, H. (2015). Late-emerging and resolving dyslexia: A follow-up study from kindergarten to grade. *Journal of Abnormal Child Psychology*, 43(7), 1389-1401. doi: 10.1007/s10802-015-0003-1

ABSTRACT

This study focuses on the stability of dyslexia status from Grade 2 to Grade 8 in four groups: (a) no dyslexia in either grade (no-dyslexia, $n = 127$); (b) no dyslexia in Grade 2 but dyslexia in Grade 8 (late-emerging, $n = 18$); (c) dyslexia in Grade 2 but not in Grade 8 (resolving, $n = 15$); and (d) dyslexia in both grades (persistent-dyslexia, $n = 22$). We examined group differences from age 3.5 to age 14 in (a) reading, vocabulary, phonology, letter knowledge, rapid naming, IQ, verbal memory; (b) familial and environmental risk and supportive factors; and (c) parental skills in reading, phonology, rapid naming, verbal memory, and vocabulary. Our findings showed group differences both in reading and cognitive skills of children as well as their parents. Parental education, book-reading frequency, and children's IQ, however, did not differentiate the groups. The children in the persistent-dyslexia group exhibited widespread language and cognitive deficits across development. Those in the resolving group had problems in language and cognitive skills only prior to school entry. In the late-emerging group, children showed clearly compromised rapid naming. Additionally, their parents had the most severe difficulties in rapid naming, a finding that suggests strong genetic liability. The findings show instability in the diagnosis of dyslexia. The members of the late-emerging group did not have a distinct early cognitive profile, so late-emerging dyslexia appears difficult to predict. Indeed, these children are at risk of not being identified and not receiving required support. This study suggests the need for continued monitoring of children's progress in literacy after the early school years.

Keywords: early identification, family risk for dyslexia, late-emerging dyslexia, reading disabilities, dyslexia

LATE-EMERGING AND RESOLVING DYSLEXIA: A FOLLOW-UP STUDY FROM AGE 3 TO 14

A considerable amount of research over the past decades has focused on reading disability (RD). This work has increased the understanding of the etiology and assessment of RD, as well as of the risk and protective factors for RD (for a recent review, see Snowling & Hulme, 2013). Most of the studies, however, have focused on the early phases of reading development or used relatively short follow-ups. The few studies that have followed reading development for several years have suggested that, although there is high stability in RD across grades at the group level (e.g., Compton, Fuchs, Fuchs, Elleman, & Gilbert, 2008; Landerl & Wimmer, 2008), there are—at least in English—also a considerable number of cases that move across the clinical threshold over time: children who are no longer affected (resolving RD; e.g., Catts, Compton, Tomblin, & Bridges, 2012) and children who do not develop RD until Grade 4 (late-emerging RD; e.g., Catts et al., 2012; Etmanskie, Partanen, & Siegel, in press; Leach, Scarborough, & Rescorla, 2003; Lipka, Lesaux, & Siegel, 2006). Little research has been conducted on the characteristics of resolving and late-emerging RD, although the proportion of late-emerging cases has been reported to be approximately 40% of all RD cases in English (Catts et al., 2012; Leach et al., 2003; Lipka et al., 2006). The scarcity of information on these resolving and late-emerging RD groups presents challenges to early identification and prevention efforts of RD.

Four previous studies have directly addressed the stability of RD classification (Catts et al., 2012; Etmanskie et al., in press; Leach et al., 2003; Lipka et al., 2006). The studies have varied in their identification of RD. Lipka et al. (2006) identified three types of RD on the basis of reading accuracy and letter knowledge in a sample of 44 children followed from Kindergarten to Grade 4. They found that 36% of all RD cases were late-emerging. Catts et al. (2012) followed 493 children from Kindergarten to Grade 10. They reported that

approximately 25% of all RD cases had late-emerging problems in decoding. Catts et al. (2012) also found a small group of children with resolving RD (5% of all RD cases). Both Lipka et al. (2006) and Catts et al. (2012) used two measures for word-reading accuracy and none for reading fluency or spelling in the identification of RD. Etmanskie et al. (in press) used a combined measure of reading accuracy and reading comprehension. Leach et al. (2003), however, used two measures for reading accuracy, two for reading fluency, and two for spelling. They tested 161 fourth- and fifth-grade students and then examined retrospectively if the children had previously been identified by teachers as having reading difficulties. Of the 66 children with RD, 21 (32%) were identified as late-emerging. Leach et al. (2003) also identified a small group of children with resolving RD (12% of all RD cases).

All of the above studies on dyslexia stability were conducted in English. In the present study, we examined the stability of RD in Finnish, which is among the most transparent orthographies. Transparency of orthography has been found to have important effects on the development of reading skills and thus the findings in English-speaking samples may not be applicable to other, more transparent orthographies. In transparent orthographies, such as those of German or Finnish, letter–sound connections are much more consistent (Seymour, Aro, & Erskine, 2003) and children learn to decode much more quickly in these languages than children do in English (Aro & Wimmer, 2003). In Finland, for example, there are only 23 fully consistent letter–sound connections to learn for accurate decoding of all Finnish words and most children are accurate decoders after just a few weeks in school. Research on the different developmental trajectories of the aforementioned RD groups in the context of transparent orthographies is, however, completely lacking although there is correlative evidence of strong stability in reading development also in transparent languages (e.g., de Jong & van der Leij, 2002; Landerl & Wimmer, 2008; Parrila, Aunola, Leskinen, Nurmi, & Kirby, 2005). On the other hand, the correlation coefficients are still far from unity,

suggesting that different developmental trajectories may also exist in transparent orthographies.

In transparent orthographies the main characteristic of readers with RD is slow reading (e.g., de Jong & van der Leij, 2003; Landerl & Wimmer, 2008; Landerl, Wimmer, & Frith, 1997; Zoccolotti et al., 2005) while reading accuracy has been shown to be an easy skill to acquire (Aro & Wimmer, 2003; Seymour, Aro, & Erskine, 2003). Therefore, unlike previous studies on RD stability in English, we focus on reading fluency. We use multiple measures for reading speed, which are age-appropriate but otherwise very similar across ages, to ensure dyslexia criteria as consistently as possible across ages. We categorize children as (a) no dyslexia, (b) late-emerging dyslexia, (c) resolving dyslexia, and (d) persistent dyslexia based on their reading speeds in Grade 2 and in Grade 8. In addition, we validate the differences between groups by comparing their reading speed performance in tasks that were not used in the RD grouping, as well as by comparing the reading speed of the follow-up sample to the level of their classmates. Finally, we will test how many of the children are expected to change dyslexia group from Grade 2 to Grade 8 randomly due to limits of measurement reliability in order to see if the changes in dyslexia groups are truly developmental. Although we do acknowledge that reading skill is continuous, the categorical approach has clinical relevance and allows comparisons to the previous research conducted in English.

In terms of identification of risk and protective factors, as well as early identification of children at risk for dyslexia, an examination of the differences among the developmental RD groups is important. Differences in cognitive profiles is one potential explanation for the differences in the emergence of RD. Children with poor early reading skills that resolve later on may have different cognitive vulnerabilities and strengths than children with late emerging or persistent dyslexia. In the early phases of reading development letter knowledge and phonological skills seem to be particularly important, as basic decoding requires solid

knowledge of letters, sounds, and their connections (e.g. Georgiou, Parrila, & Papadopoulos, 2008; Puolakanaho et al., 2007). Later on, decoding becomes automatized in typically developing children. However, reading continues to be slow and laborious in poor readers, partly because many words are still decoded letter-by-letter (Eklund, Torppa, Aro, Leppänen, & Lyytinen, 2015; Marinus & de Jong, 2010). Later reading fluency is predicted by rapid naming (Lervåg & Hulme, 2009; Torppa et al., 2012, van Bergen, van der Leij, & de Jong, 2014). Relating this to the identified groups, the persistent group is expected to show problems in all reading related cognitive skills. Yet our focus is particularly on the cognitive profiles of the two unstable groups. On the one hand, the resolving group may lag behind in acquiring phonological skills and grapheme-phoneme knowledge. On the other hand, the late-emerging group can be expected to have problems particularly in rapid naming, which would not be apparent until the demands on fluency increase.

Two previous studies on the stability of dyslexia diagnosis have also compared the groups in terms of their cognitive skills (Catts et al., 2012; Lipka et al., 2006). Catts et al. (2012) concluded that all of the RD groups showed Kindergarten-age cognitive deficits in comparison to typical readers, but the groups did not differ from each other. The late-emerging group with problems in word reading alone had difficulties particularly in phonological awareness and sentence repetition. The resolving group showed problems of phonological awareness and letter identification, but since the size of this group was small ($n = 11$) and consisted of mixed cases with difficulties either in reading comprehension, in word reading, or in both, the findings call for replication. Similarly to Catts et al. (2012), Lipka et al. (2006) reported that their late-emerging group had difficulties in phonological awareness. It was suggested that the children's phonological skills were sufficient for the early grades, but that they started to fall behind when cognitive demands in reading increased. In the present study we include, in addition to phonological skills, several other skills that have been

shown to be closely linked to reading development: letter knowledge, rapid naming, verbal short-term memory, and vocabulary (e.g., Puolakanaho, et al., 2007; Snowling, Gallagher, & Frith, 2003; van Bergen, de Jong, Maassen, & van der Leij, 2014). Unlike previous studies, we report performance on these skills both prior to and following school entry.

In addition to cognitive skills, other risk or supportive factors may explain developmental differences among the groups. One such factor is family risk for dyslexia, which was not examined in the previous studies on the stability of RD. The risk for dyslexia has been reported to range from fourfold (Puolakanaho, et al., 2007) to tenfold (van Bergen, de Jong, Plakas, Maassen, & van der Leij, 2012) for children with family risk compared to children without such risk. Family risk has predicted children's reading development over and above children's skills in the key cognitive precursors, such as phonological awareness, rapid naming, and letter knowledge (Puolakanaho et al., 2007; Torppa, Eklund, van Bergen, & Lyytinen, 2011). Furthermore, studies predicting children's skills with parent's skills have suggested that specific parental skills may be informative in assessing children's liability for dyslexia beyond their own cognitive development (Torppa et al., 2011; van Bergen et al., 2014a).

A third factor that could explain the differential developmental trajectories is the amount of environmental support. Leach et al. (2003) compared the groups in terms of print exposure, but their study did not find differences between the early-emerging and late-emerging groups. However, they used an author-recognition test, which is an indirect measure of children's print exposure. It remains possible, therefore, that a more direct evaluation of the amount of reading activities would find differences between the RD groups. In the present study we examine the amount of book reading children do with their parents and the amount they do alone. We also examine group differences in parental education. Finally, we examine gender distributions among the groups. Both gender and print exposure

comparisons were motivated by consistent findings of a gender gap in literacy, which is often attributed to fewer reading activities among boys (see OECD, 2010a, b).

This paper examines the following research questions: What is the instability of dyslexia between Grade 2 and 8? Do children change dyslexia status more often than unreliability of diagnostic tests predicts? Do the four groups differ in (a) the development of reading speed, (b) the development of language and cognitive skills, (c) the amount of book reading, (d) gender, or (e) parental education and reading(-related) skills?

METHOD

Participants

All children ($n = 182$)¹ were participants of the Jyväskylä Longitudinal Study of Dyslexia (JLD) (see Lyytinen, et al., 2008), originally selected for one of two samples: those with family risk for dyslexia or those without it. Children at risk ($n = 101$) had a parent and one or more other close family members with dyslexia. The parents' dyslexia status was confirmed through an extensive test battery (see Leinonen, et al., 2001). All children spoke Finnish as their native language and had no mental, physical, or sensory impairments. None of the children had a standard score below 80 in both performance and verbal IQ assessed in Grade 2 (WISC-III-R; Wechsler, 1991). There were 86 girls and 96 boys in the sample. In addition to the follow-up sample, their classmates' reading skills were assessed in Grade 2 ($n = 1356$), Grade 3 ($n = 2575$), and Grade 7 ($n = 1451$). The classmates' data provided a reference point for typical development. All children attended mainstream public schools following the national curriculum. JLD has received ethical consent from University of Jyväskylä ethical board.

Measures

1) Of the 200 children originally screened, 18 (of whom 7 had a familial risk for dyslexia) did not wish to take part in the eighth-grade assessments.

Children's cognitive and literacy skills. The children's cognitive and literacy skills were assessed individually by trained testers prior to school entry (from age 3.5 to 6.5) and in Grades 2, 3, and 8. Children in Finland enter Grade 1 in the fall of the year they turn 7 years.

Vocabulary. At age 3.5 and 5.5, the Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1983) was used. The Finnish translation of the BNT (Laine et al., 1993; Laine, Koivuselkä-Sallinen, Hänninen, & Niemi, 1997) contains 60 pictured items which the child is asked to name. Testing is continued until six consecutive errors are incurred. The score is based on the total number of items that are spontaneously correct plus the number of items correctly identified following a semantic stimulus cue (e.g. violin – an instrument, tennis racket – you play a game with it). In Grade 2, the WISC-III (Wechsler, 1991) was used.

Memory. Verbal short-term memory was assessed at age 6.5 and in Grades 3 and 8 with a forward digit span test. The measure was the number of correctly repeated number sequences of 12 items.

Phonological awareness. At ages 4.5, 5.5, and 6.5, phonological awareness was measured with a composite mean of z scores from four tasks: first phoneme identification, first phoneme production, segment identification, and synthesis (Torppa, et al., 2007). Cronbach's alphas were .71 at age 4.5, .58 at age 5.5, and .85 at age 6.5. In Grades 3 and 8, the common unit task was used: the task was to repeat aloud a sound that was common to two different pseudowords presented via earphones (Torppa, Georgiou, Salmi, Eklund, & Lyytinen, 2012). The score was the number of correct responses (phoneme or letter name) out of 15 items. Cronbach's alpha was .81 in Grade 3 and .85 in Grade 8.

Rapid naming. Children were asked to name as rapidly as possible, a matrix of 30 (age 5.5. years) or 50 objects ((age 6.5, and in Grades 2, 3, and 8; Denckla & Rudel, 1974)) made up from five different pictures of objects: a car, a house, a fish, a pencil and a ball. All

the Finnish names for these objects are two-syllabic high-frequency words. Total naming time (in seconds) was used as the score.

Letter knowledge. All 29 lowercase letters (23 typically used and 6 for the rare loan words) in the Finnish alphabet were presented at ages 4.5, 5.5, and 6.5. The measure was the number of correctly named letters. Cronbach's alphas were .83 at age 4.5, .88 at age 5.5, and .93 at age 6.5.

IQ (Grade 2). Four performance-quotient subtests (Picture Completion, Block Design, Object Assembly, and Coding) and five verbal-performance subtests (Similarities, Vocabulary, Comprehension, Series of numbers, and Arithmetic) of the WISC-III-R were used. The estimate of the IQ was calculated according to the manual. The Cronbach's alpha for the composite of the subtests was .70.

Children's reading fluency. Reading fluency was assessed individually in Grades 2, 3, and 8 and in groups at school in Grades 2, 3, and 7. The Cronbach's alphas for the composites of reading fluency were .92 in Grade 2, .86 in Grade 3, and .90 in Grade 8.

Word-list reading (Grades 2, 3, and 8). In the Lukilasse nationally standardized reading test (Häyrynen, Serenius-Sirve, & Korkman, 1999), participants had 2 min (Grades 2 and 3) or 1 min (Grade 8) to read aloud as many words as possible from a 90-item (Grade 2) or 105-item (Grades 3 and 8) list. The measure of the word-list reading speed was the number of correctly read words within the time limit. The inter-rater reliability was .99.

Text reading (Grades 2, 3, and 8). Age-appropriate ordinary texts were selected with lengths of 124, 189, and 204 words (for Grades 2, 3, and 8, respectively). Total reading time was the measurement of text reading speed.

Pseudoword text reading (Grades 2, 3, and 8). Children read aloud a short text made up of 19 (Grade 2) or 38 pseudowords (Grades 3 and 8). The words and structure of the

sentences resembled real Finnish in form but had no meaning. Total reading time was the measure of pseudoword reading speed.

Wordchains (Grades 2, 3, and 7). In Grades 2 and 3 the test included 79 wordchains each containing 2–4 words, and in Grade 7 it consisted of 75 wordchains each containing 4 words. The task was administered in a group context in classrooms. The child's task was to scan and mark with a pencil the boundaries in the chain where one word ends and another starts. The number of correct answers during the time limit of 2 min (Grades 2 and 3) or 3 min and 30 s (Grade 7) was used as a measure of reading speed.

Print exposure. Print exposure was assessed via parental questionnaires on the amount of book reading (Grades 2, 3, and 7) and through self-reports (Grade 7). Prior to school entry at ages 4, 5, and 6, the amount of book reading was assessed as the amount of shared book reading with parents. To produce a composite score of shared reading, we obtained parental reports of both frequency and time spent on children's reading activities in the home. Two items assessed the frequency: How often (a) the mother reads with the child, and (b) the father reads with the child. Two items covered the amount of time spent with print materials: (a) the typical duration of a reading episode (i.e., the child reads with an adult), and (b) the total time per day the child spends reading a book with an adult. Shared reading composites were derived by calculating the mean of these four item scores. In Grades 2 and 3, a composite score was derived for two items pertaining to independent book reading: (a) frequency of reading alone, and (b) the typical duration of a reading episode. Parents responded to the first item using a five-point scale (1 = not at all/seldom ... 5 = several times a day) and to the second item using a three-point scale (1 = less than 15 min/episode... 3 = longer than 45 min/episode). In Grade 7, print exposure was based on both the child's self-report and a parental report. The questions and scales were the same as in Grades 2 and

3, a total of four items. Cronbach's alpha was .79, .84, and .84 for Grades 2, 3, and 7, respectively.

Dyslexia Criteria in Grades 2 and 8. Dyslexia criteria were based on the following tasks: (a) word list reading, (b) text reading, and (c) pseudoword text reading. First, a cut-off criterion for deficient performance was defined for each measure, using the 10th percentile in the distribution of the children without family risk ($n = 81$). Subsequently, a child was considered to have dyslexia if the child scored below the criterion in at least two out of three measures of reading speed. In comparison to the larger samples with classmates ($n = 1386$ and $n = 1489$ in Grade 2 and 7, respectively) the mean reading skill of children having dyslexia were at the level of the 8th and the 6th percentile in Grade 2 and 7, respectively.

Parental assessment. The literacy skills of the parents were assessed before the child's birth. In the present study we included text reading speed because it resembles the children's tasks. When the children were between ages 3 and 6, we invited the parents for reassessments to measure their reading-related cognitive skills. Because we were not able to reassess all parents, the sample size for cognitive measures is somewhat lower than for reading speed (i.e., $n = 74$ vs. $n = 100$ for the at-risk group parents and $n = 45$ vs. $n = 81$ for the control parents with typical reading skills). Comparisons of the attendees and non-attendees revealed that the educational level and age of the parents was not different. There were, however, differences in parental reading skills: the average reading level of the parents who decided to attend reassessments was somewhat lower than that of the non-attendees. For one of the parents in the family risk group, the text reading task was very difficult and testing was discontinued.

Text-reading fluency. Parents read aloud two passages (218 and 128 words, respectively) as fluently and accurately as possible. A measure of reading fluency was the average reading time for the two texts. Cronbach's alpha was .96.

Phoneme deletion. Parents pronounced a given word without the second phoneme. The task included 16 words (e.g., kaupunki ‘city’ became kupunki) of 4 to 10 letters with 2 to 4 syllables. Deletion of the second phoneme yielded a pseudoword. Stimuli were presented via headphones. A new stimulus was presented after a response or after a 20-second period of silence. The number of correct responses was calculated.

Rapid Naming. On each of three tasks, participants named, as rapidly as possible, a matrix of 50 items comprising objects, digits, or a mixture of digits, objects, and letters. In the parental assessment, stimuli were presented on a computer screen. Total naming time (in seconds) was used as the score. A mean composite score of the three standardized rapid naming scores was calculated for the analyses. Cronbach’s alpha for the rapid naming composite was .86.

Verbal short-term memory. In the digit span subtest of WAIS-III (Wechsler, 1991), participants repeated strings of digits, increasing in length, in both the forward and reverse directions. Two sets of items, one for forward, the other for reverse, were used. Scaled scores were derived from the manual.

Vocabulary. In the vocabulary subtest of WAIS-III (Wechsler, 1991), participants defined 35 words in their own words. Scaled scores were derived from the manual.

RESULTS

We first classified children in four groups, according to their dyslexia status in Grades 2 and 8: (a) no dyslexia in either grade (no-dyslexia, $n = 127$); (b) no dyslexia in Grade 2 but dyslexia in Grade 8 (late-emerging, $n = 18$); (c) dyslexia in Grade 2 but not in Grade 8 (resolving, $n = 15$); and (d) dyslexia in both grades (persistent-dyslexia, $n = 22$). Because we classify children using cut-offs at two time points, which are always somewhat arbitrary (e.g. Francis et al., 2005), and because individual test scores are never 100% reliable, we first examined how many of the Grade 2 children would be expected to have changed group by

Grade 8 just by chance. To do this we did a simulation study using the reliability of the three reading tasks (measured by test-retest correlations between Grades 2 and 3). Our 1-year test-retest correlations reflect both measurement unreliability and trait stability. Therefore, they estimate reliability conservatively. Reading measures between Grade 2 and 3 correlated .80 for word list reading, .66 for pseudoword text reading, and .87 for text reading. We decided to use a test-retest correlation estimate of .80 (average of the test-retest correlations of the three tasks) and trait true score stability was set to be 1. In the simulation study we set the number of cases to be 100,000 and examined how many children would be observed to change their group if we assume that there is no true score changes but all changes are seen only in the observed scores (being thus random changes due to unreliability of measurement). In the simulation the same identification procedure with three measures and 10% cut-off was used as was used for the identification in the other analyses. The results showed that 5.6% , which would correspond to 10 children in our sample, change their group just because of the unreliability of the measures between Grades 2 and 8. As 18% of our sample (33 children) changed their diagnostic group, we conclude that, even with the conservative reliability estimate, the instability of dyslexia is not fully due to random changes across cut-offs.

Group Differences in Parental Education, Gender, and Family Risk

There were no differences between groups in IQ or parents' education (see Table 1). Children's gender was unevenly distributed in the groups, $\chi^2(3) = 11.06, p < .05$: there were more boys than expected in the late-emerging group (adjusted standardized residual = 2.5) and more girls in the resolving group (adjusted standardized residual = 2.4). In addition, family risk for dyslexia was unevenly distributed in the groups, $\chi^2(3) = 19.54, p < .001$: at-risk children were overrepresented in the persistent and late-emerging groups (adjusted

standardized residual = 2.6 and 2.5, respectively) and underrepresented in the no-dyslexia group (adjusted standardized residual = -4.4).

Group Differences in Reading Speed Development

In order to examine if the groups were different in reading speed development, we conducted group comparisons using both the individually administered dyslexia criterion tasks (in Grades 2, 3, and 8) and group administered tasks (in Grades 2, 3, and 7) that were not part of the dyslexia criterion. The group tasks were also administered to children's classmates in order to obtain a reference sample. Group comparisons were conducted with one-way ANOVAs (see Figure 1, Table 2). Note that in Figure 1, there are two different distributions underlying the standardization. In the left panel for the individually administered tasks, the standardization is based on the not-at-risk group's distribution. In the right panel for the group-administered reading tasks, the standardization is based on the larger sample, which includes the classmates of the follow-up group as well. Effect sizes (Cohen's *d* computed using pooled standard deviation) are reported in Supplementary table 1.

All dyslexia groups showed poorer performance than the no-dyslexia group across measures and time-points. The persistent group stayed at the lowest level throughout the whole follow-up period. The late-emerging group read also somewhat slower than the no-dyslexia group in Grades 2 and 3 and showed a descending trajectory in the follow-up to Grades 7 and 8. The resolving group, on the other hand, did not differ in reading speed from the persistent group in Grades 2 and 3, but showed fast development in the following years. We also conducted a follow-up analysis on the developmental differences in reading speed among the groups with mixed ANOVA where we entered grade as a within-subjects factor and dyslexia group as a between-subjects factor. There was a significant Grade x Group interaction effect, $F(3, 174) = 18.50, p < .001, \eta_p^2 = .24$, which indicates group differences in the rate of reading speed development between Grades 2 and 8. The clearest difference was

between the resolving and late-emerging group: the resolving group made progress in reading speed more quickly than the late-emerging group did.

Group Differences in Cognitive Skills

Next we compared the groups in terms of cognitive skills with one-way ANOVAs (see Table 3 and Supplementary Table 2, Figure 2). The persistent-dyslexia group showed poor performance in almost all assessed cognitive skills. They performed below the no-dyslexia group in all measures except vocabulary, and phonological awareness at ages 4.5 and 5.5. The strongest effect sizes for the difference between the persistent and no-dyslexia groups were found in phonological awareness from age 6.5 onwards, in rapid naming, and in letter knowledge.

The late-emerging group had problems most clearly in rapid naming speed. They performed below the no-dyslexia group in all assessments of rapid naming and also in Grade 3 verbal short-term memory. In addition, they tended to perform below the no-dyslexia group in the other cognitive skills as well (except for vocabulary), although these medium effect size differences did not reach significance. The late-emerging group and the persistent-dyslexia group differed significantly only in phonological awareness (Grade 8), although medium to strong effect sizes in favor of the late-emerging group emerged for verbal short-term memory (age 6.5) and for letter knowledge (ages 5.5 and 6.5).

The resolving group had problems in cognitive skills but only prior to school entry: they performed below the no-dyslexia group in vocabulary (ages 3.5 and 5), phonological awareness (age 6.5), rapid naming (age 5.5), and letter knowledge (all occasions). In addition, medium effect sizes emerged for all cognitive skills assessed prior to school entry in favor of the no-dyslexia group as well as for verbal short-term memory in all time-points. The resolving group did not differ from the persistent group significantly except for phonological awareness in Grades 3 and 8. However, medium effect sizes in favor of the resolving group

emerged also in rapid naming (Grades 3 and 8). On the other hand, medium effect sizes in favor of the persistent group were found in vocabulary (ages 3.5 and 5). The resolving group and late-emerging group did not differ significantly from each other in any of the cognitive measures. However, there were medium effect sizes in favor of the late-emerging group before school-age in vocabulary (age 5), verbal short term memory (age 6.5) and letter knowledge (age 6.5). On the other hand, at school-age medium effect sizes were found in favor of the resolving group in phonological awareness (Grade 3) and in rapid naming (Grade 8).

Group Differences in Book Reading

Finally, the groups were compared in the amount of book reading (the amount of shared book reading with a parent prior to school entry and the amount of book reading alone after school entry). According to the pairwise comparisons, the groups did not differ in the amount of book reading although the F-test in Grade 3 was significant. The comparisons of the effects sizes suggested that the parents of the resolving group tended to read less to their 4-year olds than the late-emerging and persistent-dyslexia groups did (medium effect sizes). In addition, an effect size of .71 emerged for the comparison of the no-dyslexia group and resolving group in Grade 3, a result that indicates that the no-dyslexia children spent more time reading books than the resolving children.

Group Differences in Parental Skills

For the group comparisons in parental skills, five groups were compared instead of four because the no-dyslexia group was split into two groups: children with family risk for dyslexia and children without family risk for dyslexia. This division was made in order to provide a more detailed examination of the parental skill differences in these two separate samples. There were 3–4 children without family risk in each of the dyslexia groups whose data was omitted from this comparison resulting in five groups: not-at-risk & no-dyslexia (*n*

= 70), at-risk & no-dyslexia ($n = 57$), at-risk & late-emerging dyslexia ($n = 15$), at-risk & resolving dyslexia ($n = 11$), at-risk & persistent-dyslexia ($n = 18$). The group comparisons are reported in Table 4, Supplementary table 3, and Figure 3.

The group comparisons revealed that parents in the not-at-risk & no-dyslexia group performed significantly better than parents in all other groups in reading speed and in phoneme deletion. These parents also performed better in the verbal short-term memory task than parents in all other groups did except for those in the at-risk & resolving group. There were no significant differences in parental vocabulary although large effect sizes indicated that the parents in the at-risk persistent groups tended to have poorer vocabulary than in those in the not-at-risk & no dyslexia group did. The group comparisons in parents' rapid naming showed interesting differences between the family risk groups: the parents of the at-risk & late-emerging and the at-risk & persistent dyslexia groups were slow at rapid naming whereas the parents of the at-risk & resolving and the at-risk & no-dyslexia groups did not differ from the not-at-risk & no-dyslexia group's parents in rapid naming. The effects sizes confirmed that the parents of at-risk & late-emerging and at-risk & persistent dyslexia groups had slow rapid naming speed when they were compared with both no-dyslexia groups and with the at-risk & resolving group.

DISCUSSION

The current study investigated the stability of dyslexia in a prospective study from Kindergarten to Grade 8 in the context of a transparent orthography (Finnish). The children, half of whom had a familial risk for dyslexia, were categorized as having or not having dyslexia based on reading fluency measures in Grades 2 and 8. This yielded three groups of children with reading problems at some point, referred to as resolving, late-emerging, and persistent dyslexia. The three groups were compared with each other and with a group without dyslexia in six measures: reading fluency, language and cognitive skills, parental

reading fluency and cognitive skills, parental education, and the amount of book reading. The group comparisons revealed differences both in the children's reading-related cognitive development and in a set of similar skills measured in their parents. Parental education, book reading frequency, and children's IQ, on the other hand, did not differentiate the groups.

Our findings indicated that reading status was not stable, because less than half of the RD children met the dyslexia criteria in both Grades 2 and 8. In fact, of the 55 children identified as having RD at some point, 15 (27%) met the dyslexia criteria in Grade 2 only (resolving), and 18 (33%) only in Grade 8 (late-emerging). The investigation of reading fluency using the group-administered tasks that were external to the dyslexia criteria, validated the groups. The previous longitudinal studies examining the stability of dyslexia were conducted in English (Catts et al., 2012; Etmanskie et al., in press; Leach et al., 2003; Lipka et al., 2006). In spite of differences in orthographic complexity between this study and previous ones, we found roughly the same proportion of children with late-emerging RD as Catts et al. (2012) and Leach et al. (2003). Nevertheless, the proportion of the resolving group was twice as large than the group in Leach et al. (2003) and four times as large than what was found in Catts et al. (2012). Such cases were not reported in Lipka et al. (2006).

The letter–sound connections in Finnish, which are easier to learn than in English, could enable more children to catch up despite early cognitive difficulties. This suggestion is supported by cross-linguistic comparisons which have shown that dyslexic children's reading is less severely impaired in low-complexity orthographies than in high-complexity ones (Landerl, et al., 2012). In addition, the use of reading speed as a measure of RD and the longer follow-up period may explain why the proportion of resolving RD is higher in the present study. Although typical Finnish children are fluent readers already in the spring of Grade 2, reading fluency continues to develop rapidly. Grade 2 spring is an interesting assessment time because at this point in Finnish schools a pedagogical shift occurs: starting

from Grade 3 the emphasis on learning to read changes into learning by reading and children are expected to read more and the demands for reading speed increase. It seems that for some children, however, the development of reading fluency takes longer but that they can catch up later on.

The process of identifying these groups raises a number of questions: Can these groups of children be identified early on? What intrinsic or extrinsic factors can help children to overcome reading impairments? What factors cause them to succumb to reading impairments later on? The examination of the cognitive differences between the groups showed, in line with the body of literature on early cognitive precursors of dyslexia (e.g., Puolakanaho et al., 2007; Snowling et al., 2003; van Bergen et al., 2014b), the close link of rapid naming, letter knowledge, and phonological awareness to RD. The findings for verbal short-term memory and vocabulary were not as consistent, because all RD groups showed moderate difficulty in verbal short-term memory tasks and only the resolving group had clear early vocabulary difficulties. The cognitive difficulties were limited to skills closely linked to reading. There were indeed no group differences in IQ.

The persistent-dyslexia group had early and persisting deficits across the cognitive foundations of reading, as expected. Their performance in early phonological awareness and expressive vocabulary, however, was not significantly poorer than in the typically reading group. This result, which contradicts previous studies (Catts et al., 2012; Lipka et al., 2006) on the stability of RD, can be explained by differences in the orthography and in the RD classification criteria. The role of phonological awareness in transparent orthographies has been shown to be limited to the very beginning of reading acquisition and particularly to reading accuracy (e.g. de Jong & van der Leij, 2002; Landerl & Wimmer, 2008). The present study adopted RD criteria that uses reading fluency measures because the reading accuracy approaches a ceiling even with nonword reading measures in Grade 2. The studies conducted

in English, however, used mainly reading accuracy measures (e.g., Lipka et al., 2006) or reading accuracy and comprehension measures (e.g., Etmanski et al., in press) in their RD criteria.

The late-emerging group differed significantly from the no-dyslexia group especially in rapid naming prior to and after school entry. This finding was expected because rapid naming has been shown to be a strong predictor of reading speed measures (e.g., Puolakanaho et al., 2007; van Bergen et al., 2014a). The skills of the late-emerging group seemed to be sufficient for the early grades, but not for reaching the typical level of fluency in reading in later grades. This finding is in accordance with the idea that the major bottleneck in reading development in Finnish is in reading speed and rapid naming whereas in the case of English the development of reading accuracy may also be problematic and is linked to phonological skills. Some of the previous studies on dyslexia stability in English (e.g. Catts et al., 2012; Etmanskie et al., in press) included in their assessment reading comprehension. For comparison, we additionally looked at reading comprehension in the late-emerging group in Grades 2, 3, and 9. Their comprehension skills were age appropriate, which is in line with previous findings (e.g. Torppa et al., 2007) that in Finnish it is possible to obtain adequate reading comprehension, despite slow reading.

The resolving group showed difficulties in phonological awareness, letter identification, rapid naming, and vocabulary prior to school entry. Surprisingly, the resolving group had the lowest level of expressive vocabulary. By school age, however, the cognitive differences between the resolving and no-dyslexia groups had disappeared, suggesting that these children suffered from a developmental delay rather than from permanent cognitive deficits. Another explanation for the fast catch-up of the resolving group may be that school entry meant a clear improvement in the environmental support for reading-related skills. Although our measures of parental education and the reported amount of shared reading did

not show significant group differences, more sensitive measures of the home environment might have done so.

Comparisons between the groups in terms of parental skills also revealed interesting differences. First, as expected, children with a familial risk for dyslexia were overrepresented in the dyslexia groups. Second, parental skills were different in the groups. Interestingly, the parents of the late-emerging and persistent group had slower naming rates than the parents of the other groups, a result which matches the findings at the child level. This finding suggests that the late-emerging and persistent group had stronger vulnerability for developing difficulties in rapid naming and reading fluency. These findings support previous studies that show that parental rapid naming is predictive of their offspring's naming and reading fluency (Torppa et al., 2011; van Bergen et al., 2014b). It thus supports the notion that parental skills are informative of their offspring's liability for dyslexia (van Bergen et al., 2014b).

Finally, the gender difference between the late-emerging (22% girls) and resolving (80% girls) groups was striking. This finding is in line with the evidence that girls outperform boys in reading in upper grades (OECD, 2010a, b) and that fewer girls have reading disabilities (Rutter et al., 2004). In the present data, in Grade 2 dyslexia was as prevalent among boys and as it was among girls, whereas in grade 8, 65% of the dyslexic adolescents were male. One explanation is that girls are more motivated to do schoolwork (Li & Lerner, 2011) and to read. Our related finding of no group differences in the amount of book reading, however, does not support the link between book reading and skill development. However, a more comprehensive measurement of print exposure, also including digital reading and school engagement, might show different results.

It has been proposed that in the diagnosis of specific learning disabilities (DSM-5) (see Tannock, 2013), diagnostic criteria should include the early onset of symptoms of the disability. The late-emerging group is interesting in relation to this proposal, because they do

not meet this criterion. As a result, they did not fulfill the proposed dyslexia criteria, despite the observation that their group mean in Grade 8 reading fluency was two standard deviations below those of the unaffected adolescents. Tannock (2013) states that the symptoms “may not become fully manifest until the learning demands exceed the individual’s limited capacities” (p.19). Based on the current findings, the early symptoms of the late-emerging group are mild and may not be evident in the early grades, even if reading is assessed more than once (e.g., in Grades 2 and 3). However, slow naming speed in children and their parents seems to be a warning sign that reading speed may develop slower later on.

There are certain limitations in this work that need to be considered. First, it should be noted that our findings call for replication, because the late-emerging and resolving groups were rather small. Second, because the data came from a family risk study (see Lyytinen et al., 2008), it includes a higher prevalence of dyslexia than expected in the general population. It is therefore not suited for estimating the prevalence of persistent, resolving, and late-emerging dyslexia in population. Third, the instability of the RD definition is partially attributable to the use of a categorical approach (see Francis et al., 2005). However, our simulations showed that only 5.6% of the changes are due to random changes across cut-off criterion due to unreliability of the measures. It should be noted that the clinical question regarding the stability of dyslexia status supports the use of a categorical approach. Additionally, adopting such an approach allowed comparisons with previous investigations. It should also be noted that there are different ways of defining categories. The low achievement approach we used seems to be one of the most stable and reliable ones, although the reliability of the diagnostic tests used is also critical (Brown Waesche, Schatsneider, Maner, Ahmed, & Wagner, 2011). Our reading fluency tests showed high reliability and the group differences were also validated with external reading speed measures.

In conclusion, even in a language environment where children read very accurately after two years of reading instruction, we found that reading status was not yet stable at this age. This raises several clinical implications. First, it is important to continue following children's literacy development beyond the early grades. Second, support needs to be provided not only for those who receive an early diagnosis, but also for those who begin lagging behind later in their development. Only continuous follow-ups can detect children who fall behind later on. If an official diagnosis is needed to access extra support, which is the case in several countries, children with late-emerging dyslexia will be deprived of the intervention and adaptations they need.

ACKNOWLEDGEMENTS

This research has been supported by several funding agencies over the years. The current funding comes from the Academy of Finland (projects #264264 and #276239). Elsjé van Bergen's work was supported by a Rubicon Fellowship (446-12-005) from the Netherlands Organisation for Scientific Research (NWO) and a Junior Research Fellowship at Oriel College, Oxford. There are no conflicts of interest regarding this work. Funding sources have had no role in the collection, analysis, and interpretation of data; in the writing of the report or in the decision to submit the paper for publication. We want to express our gratitude to all the families, children, teachers, schools, and researchers who have participated in this study. Many thanks also to Professor Asko Tolvanen for statistical assistance and Matthew Wuethrich for the language proofing of this manuscript.

REFERENCES

- Aro, M., & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics*, *24*, 621-635. doi: 10.1017.S0142716403000316
- Brown Waechse, J.S., Schatshneider, C., Maner, J.K., Ahmed, Y., & Wagner, R.K. (2011). Examining agreement and longitudinal stability among traditional and response-to-intervention-based definitions of reading disability using the affected-status agreement statistic. *Journal of Learning Disabilities*, *44*, 296-307. doi: 10.1177/0022219410392048
- Catts, H. W., Compton, D., Tomblin, J. B., & Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology*, *105*, 166-181. doi: 10.1037/a0025323
- Compton, D. L., Fuchs, D., Fuchs, L. S., Elleman, A. M., & Gilbert, J. K. (2008). Tracking children who fly below the radar: latent transition modeling of students with late-emerging reading disability. *Learning and Individual Differences*, *18*, 329–337. doi:10.1016/j.lindif.2008.04.003
- 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. doi: 10.1207/S1532799XSSR0601_03
- de Jong, P.F., & van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology*, *95*, 22-40. doi: 10.1037/0022-0663.95.1.22
- Denckla, M. B., & Rudel, R. G. (1974). Rapid “automatized” naming of pictured objects, colors, letters and numbers by normal children. *Cortex*, *10*, 471-479.

- Eklund, K., Torppa, M., Aro, M., Leppänen, P.H., & Lyytinen, H. (2015). Literacy skill development of children with familial risk for dyslexia through grades 2, 3, and 8. *Journal of Educational psychology*, *107*, 126-140. doi: 10.1037/a0037121.
- Etmanskie, J.M., Partanen, M., & Siegel, L. (in press). A longitudinal examination of the persistence of late emerging reading disabilities. *Journal of Learning Disabilities*. doi: 10.1177/0022219414522706
- Francis, D. J., Fletcher, J. M., Stuebing, K. K., Lyon, G. R., Shaywitz, B. A., and Shaywitz, S. E. (2005). Psychometric approaches to the identification of LD: IQ and achievement scores are not sufficient. *Journal of Learning Disabilities* *38*, 98–108. doi: 10.1177/00222194050380020101
- Georgiou, G., Parrila, R., & Papadopoulos, T. C. (2008). Predictors of word decoding and reading fluency in English and Greek: a cross-linguistic comparison. *Journal of Educational Psychology*, *100*, 566–580. doi: 10.1037/0022-0663.100.3.566
- Häyrinen, T., Serenius-Sirve, S., & Korkman, M. (1999). *Lukilasse*. Helsinki, Finland: Psykologien Kustannus Oy.
- Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *The Boston Naming Test* (2nd ed.). Philadelphia: Lea & Febiger.
- Laakso, M.L., Poikkeus, A.M., Eklund, K., & Lyytinen, P. (2004). Children's interest in early shared reading and its relation to later language and letter knowledge in children with and without genetic risk for dyslexia. *First Language*, *24*, 323-345. doi:10.1177/0142723704046041.
- Laine, M., Goodglass, H., Niemi, J., Koivuselkä-Sallinen, P., Tuomainen, J. & Marttila, R. (1993). Adaptation of the Boston Diagnostic Aphasia Examination and the Boston Naming Test into Finnish. *Scandinavian Journal of Logopedics & Phoniatrics*, *18*, 83–92. doi:10.3109/14015439309101353

- Laine, M., Koivuselkä-Sallinen, P., Hänninen, R. & Niemi, J. (1997). *Bostonin nimentätesti*. [Finnish translation of the Boston Naming Test] Helsinki: Psykologien Kustannus Oy.
- Lervåg, A., & Hulme, C. (2009). Rapid naming (RAN) taps a basic constraint on the development of early reading fluency. *Psychological Science, 20*, 1040-1048. doi: 10.1111/j.1467-9280.2009.02405.x
- Landerl, K., Ramus, F., Moll, K., Lyytinen, H., Leppänen, P. H. T., Lohvansuu, K., ... Schulte-Körne, G. (2012). Predictors of developmental dyslexia in European orthographies with varying complexity. *Journal of Child Psychology and Psychiatry, 54*, 686-694. doi: 10.1111/jcpp.12029
- 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., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: a German-English comparison. *Cognition, 63*, 315-334. doi: 10.1016/S0010-0277(97)00005-X
- Leach, J. M., Scarborough, H. S., & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology, 95*, 211-224. doi:10.1037/0022-0663.95.2.211.
- Leinonen, S., Müller, K., Leppänen, P. H. T., Aro, M., Ahonen, T., & Lyytinen, H. (2001). Heterogeneity in adult dyslexic readers: relating processing skills to the speed and accuracy of oral text reading. *Reading and Writing: An Interdisciplinary Journal, 14*, 265-296. doi: 10.1023/A:1011117620895
- Li, Y. & Lerner, R. M. (2011). Trajectories of school engagement during adolescence: Implications for grades, depression, delinquency, and substance use. *Developmental Psychology, 47*, 233-247. doi:10.1037/a0021307

- Lipka, O., Lesaux, N., & Siegel, L. (2006). Retrospective analyses of the reading development of Grade 4 students with reading disabilities: risk status and profiles over 5 years. *Journal of Learning Disabilities, 39*, 364–378.
doi:10.1177/00222194060390040901
- Lyytinen, H., Erskine, J., Ahonen, T., Aro, M., Eklund, K., Guttorm, T., ... Viholainen, H. (2008). Early identification and prevention of dyslexia: Results from a prospective follow-up study of children at familial risk for dyslexia. In G. Reid, F. Manis, & L. Siegel (Eds.), *The Sage handbook of dyslexia* (pp. 121–146). Thousand Oaks, CA: Sage.
- Marinus, E. & de Jong, P.F. (2010). Variability in the word reading performance of dyslexics: effects of word length and digraph presence. *Cortex, 46*, 1259-1271.
doi:10.1016/j.cortex.2010.06.005
- OECD (2010a). *PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I)*, PISA, OECD Publishing.
doi:10.1787/9789264091450-en
- OECD (2010b). *PISA 2009 Results: Learning to Learn – Student Engagement, Strategies and Practices (Volume III)*, PISA, OECD Publishing. doi:10.1787/9789264083943-en
- Parrila, R., Aunola, K., Leskinen, E., Nurmi, J.-E., & Kirby, J. R. (2005). Development of individual differences in reading: results from longitudinal studies in English and Finnish. *Journal of Educational Psychology, 97*, 299-319. doi:10.1037/0022-0663.97.3.299
- Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Leppänen, P.H.T., Poikkeus, A.-M., Tolvanen, A., Torppa, M., & Lyytinen, H. (2007). Very early phonological and language skills: estimating individual risk of reading disability. *Journal of Child Psychology and Psychiatry, 48*, 923-931. doi: 10.1111/j.1469-7610.2007.01763.x

- Reynolds, C. R. & Kamphaus, R. W. (1992). *BASC. Behavior Assessment System for Children. Manual*. Circle Pines, MN: American Guidance Services, Inc.
- Rutter, M., Caspi, A., Fergusson, D., Horwood, L. J., Goodman, R., Maughan, B., Moffitt, T. E., Meltzer, H., & Carroll, J. (2004). Sex differences in developmental reading disability: new findings from 4 epidemiological studies. *Journal of American Medical Association*, *291*, 2007-2012. doi:10.1001/jama.291.16.2007
- Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*, 143–174.
- Snowling, M. J. & Hulme, C. (2013). Children's reading impairments: from theory to practice. *Japanese Psychological Research*, *55*, 186-202. doi:10.1111/j.1468-5884.2012.00541.x
- Snowling, M. J., Gallagher, A., & Frith, U. (2003). Family risk of dyslexia is continuous: individual differences in the precursors of reading skills. *Child Development*, *74*, 358-373. doi: 10.1111/1467-8624.7402003
- Tannock, R. (2013). Rethinking ADHD and LD in DSM-5: Proposed changes in diagnostic criteria. *Journal of Learning Disabilities*, *46*, 5-25. doi:10.1177/0022219412464341
- Torppa, M., Eklund, K., van Bergen, E., Lyytinen, H. (2011). Parental literacy predicts children's literacy: a longitudinal family-risk study. *Dyslexia*, *17*, 339-355. doi: 10.1002/dys.437
- Torppa, M., Georgiou, G. K., Salmi, P., Eklund, K., & Lyytinen, H. (2012). Examining the double-deficit hypothesis in an orthographically consistent language. *Scientific Studies of Reading*, *16*, 287–315. doi:10.1080/10888438.2011.554470
- Torppa, M., Tolvanen, A., Poikkeus, A.-M., Eklund, K., Lerkkanen, M.-K., Leskinen, E., & Lyytinen, H. (2007). Reading development subtypes and their early characteristics. *Annals of Dyslexia*, *57*, 3-52. doi: 10.1007/s11881-007-0003-0

- van Bergen, E., de Jong, P.F., Maassen, B., & van der Leij, A. (2014a). The effect of parents' literacy skills and children's preliteracy skills on the risk of dyslexia. *Journal of Abnormal Child Psychology*, 42, 1187-1200. doi: 10.1007/s10802-014-9858-9
- 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 and Psychiatry*, 53, 28-36. doi:10.1111/j.1469-7610.2011.02418.x
- van Bergen, E., van der Leij, A., & de Jong, P.F. (2014b). The intergenerational multiple deficit model and the case of dyslexia. *Frontiers in Human Neuroscience*, 8, 346. doi: 10.3389/fnhum.2014.00346
- Wechsler, D. (1991). *Wechsler Intelligence Scales for Children* (3rd ed.). Sidcup, United Kingdom: The Psychological Corporation.
- Zoccolotti, P., De Luca, M., Di Pace, E., Gasperini, F., Judica, A., & Spinelli, D. (2005). Word length effect in early reading and in developmental dyslexia. *Brain and Language*, 93, 369-373. doi:10.1016/j.bandl.2004.10.010

Table 1. Characteristics of Parents and Children in the Four Groups.

	No-dyslexia		Late-emerging		Resolving		Persistent	
	<i>n</i> = 127		<i>n</i> = 18		<i>n</i> = 15		<i>n</i> = 22	
	M	SD	M	SD	M	SD	M	SD
Mother	4.52	1.38	4.59	1.70	3.92	1.24	4.12	1.50
Father	3.72	1.42	4.06	1.71	3.64	.81	3.75	1.39
<i>Children</i>								
IQ	100.74	10.93	100.83	6.04	97.47	9.10	95.00	10.37
Family risk	44.9%		83.3%		73.3%		81.8%	
Girls	49.6%		22.2%		80.0%		45.5%	

Note.

^{a)} Parental education was classified using a 7-point scale (see Laakso, Poikkeus, Eklund, and Lyytinen, 2004).

Table 2. Descriptive Statistics and Group Comparisons in Reading Speed.

	No-dyslexia		Late-emerging		Resolving		Persistent		F	df
	M	SD	M	SD	M	SD	M	SD		
<i>Reading speed, individual assessment^a</i>										
Grade 2	6.09 ¹	1.70	4.14 ²	.92	2.84 ³	.58	2.33 ³	.68	58.27***	3,174
Grade 3	6.66 ¹	1.65	4.80 ²	.67	4.56 ²	.94	3.33 ³	.91	41.82***	3,178
Grade 8	9.56 ¹	1.52	6.44 ³	.47	8.43 ²	1.17	5.67 ³	1.19	67.25***	3,178
<i>Reading speed, group assessment^b</i>										
Grade 2	.20 ¹	.80	-.31 ²	.56	-.94 ³	.43	-1.38 ³	.61	32.99***	3,161
Grade 3	.48 ¹	1.05	-.19 ²	.44	-.65 ^{2,3}	.64	-1.21 ³	.62	23.94***	3,171
Grade 7	.30 ¹	.90	-.87 ²	.49	-.56 ²	.66	-1.44 ³	.60	30.87***	3,144

^a) Reading speed, individual assessment = Arithmetic mean of the number of read letters/second in Word list, Text and Pseudoword text reading.

^b) Reading speed, group assessment = Z-score value in Wordchain task calculated with the mean and standard deviation values of the classmates (N=1537, 2757 and 1693 at Grade 2, 3, and 7, respectively).

Note. Groups with different superscripts (^{1,2,3}) differed from each other in post hoc pairwise comparisons where Dunnett T3 corrections were used (because the variances were not equal).

*** $p \leq .001$

Table 3. Descriptive Statistics and RD Group Comparisons in Language, Cognitive Skills and Print Exposure.

	No-dyslexia		Late-emerging		Resolving		Persistent		F	df
	M	SD	M	SD	M	SD	M	SD		
<i>Vocabulary</i>										
3.5 years	19.46 ¹	5.99	18.72 ^{1,2}	7.00	14.93 ²	5.16	17.60 ^{1,2}	4.83	2.93*	3,179
5.0 years	35.16 ¹	6.21	33.33 ^{1,2}	7.72	30.40 ²	7.57	34.15 ^{1,2}	4.55	2.77*	3,174
Grade 2	10.36	2.87	10.72	2.58	10.20	2.34	9.55	2.13	0.66	3,176
<i>Verbal short-term memory</i>										
6.5 years	5.50 ¹	1.27	5.56 ^{1,2}	1.20	4.60 ^{1,2}	1.12	4.67 ²	1.68	4.32**	3,177
Grade 2	6.44 ¹	1.11	5.94 ^{1,2}	.72	5.73 ^{1,2}	.96	5.73 ²	1.08	4.79**	3,175
Grade 3	7.71 ¹	1.31	6.61 ²	1.14	6.73 ^{1,2}	1.58	6.73 ²	1.72	7.13***	3,177
Grade 8	7.95 ¹	1.58	7.28 ^{1,2}	1.27	6.93 ^{1,2}	1.44	6.76 ²	1.51	5.42**	3,177
<i>Phonological awareness</i>										
4.5 years ^z	-.02	.80	-.51	.74	-.52	.94	-.46	.47	4.59**	3,179
5.5 years ^z	-.15	.79	-.42	.63	-.60	.78	-.46	1.07	2.26	3,180
6.5 years ^z	-.01 ¹	.91	-.59 ^{1,2}	.97	-.82 ²	.94	-.70 ²	.80	7.49***	3,177
Grade 3	9.54 ¹	3.43	7.28 ^{1,2,3}	3.75	9.79 ^{1,2}	3.26	5.73 ³	3.33	9.11***	3,177
Grade 8	10.20 ¹	3.43	8.61 ^{1,2}	3.15	9.07 ^{1,2}	2.55	5.95 ³	3.53	10.56***	3,178

<i>Rapid naming</i>										
5.5 years	42.41 ¹	10.35	55.46 ²	20.83	54.50 ²	20.64	53.10 ²	17.62	9.47***	3,178
6.5 years	68.86 ¹	18.55	85.19 ²	25.83	83.78 ^{1,2}	14.39	87.19 ²	32.07	8.01***	3,176
Grade 3	46.34 ¹	6.94	51.65 ²	7.47	48.79 ^{1,2}	7.77	54.72 ²	11.66	8.81***	3,177
Grade 8	36.56 ¹	5.37	41.12 ²	6.45	37.76 ^{1,2}	4.07	41.09 ²	4.84	7.35***	3,178
<i>Letter knowledge</i>										
4.5 years	9.30 ¹	7.83	5.22 ^{1,2}	6.22	3.53 ²	4.19	3.95 ²	4.44	6.35**	3,178
5.5 years	15.04 ¹	7.12	11.12 ^{1,2}	8.35	7.73 ²	6.58	6.85 ²	5.81	11.71***	3,178
6.5 years	18.28 ¹	7.74	14.11 ^{1,2}	7.58	9.73 ²	6.40	10.55 ²	5.39	11.49***	3,176
<i>Amount of book reading</i>										
4.0 years	9.49	2.31	10.31	2.70	8.86	1.46	9.88	1.76	1.21	3,167
5.0 years	9.41	1.99	9.38	2.53	8.71	2.27	9.47	2.21	.49	3,167
6.0 years	9.42	2.61	10.34	2.73	8.97	2.61	9.31	1.78	1.07	3,167
Grade 2	1.97	.65	1.96	.54	1.82	.58	1.91	.42	.27	3,155
Grade 3	2.23	.61	2.00	.45	1.77	.68	1.97	.50	3.66*	3,157
Grade 7	2.89	1.02	2.70	.85	3.04	1.16	2.72	.78	.44	3,149

Note. Groups with different superscripts (^{1,2,3}) differed from each other in post-hoc pairwise comparisons, with either Bonferroni or Dunnett T3 corrections, depending on equality of the variances. z = a composite mean of standardized scores

* $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$

Table 4

Descriptive Statistics and RD Group comparisons of Parental Measures

	Not at-risk		At-risk		At-risk		At-risk		At-risk		<i>F</i>	<i>df</i>
	No-dyslexia		No-dyslexia		Late-emerging		Resolving		Persistent			
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Text-reading fluency (s.)	29.62 ¹	2.70	40.55 ²	10.13	42.70 ²	8.29	39.61 ²	9.15	40.85 ²	7.56	31.65***	4,181
Phoneme deletion	11.43 ¹	.99	9.21 ²	2.10	8.70 ²	2.26	9.44 ²	1.67	8.73 ²	2.37	10.57***	4,117
Rapid naming(s.)	39.16 ¹	8.46	43.32 ^{1,2}	8.15	53.44 ³	8.80	41.36 ^{1,2}	5.63	49.28 ^{2,3}	9.58	7.58***	4,118
Verbal STM	14.73 ¹	3.68	11.57 ²	2.55	10.80 ²	1.87	12.78 ^{1,2}	3.07	10.37 ²	3.66	7.41***	4,116
Vocabulary	54.24	5.63	49.72	7.52	49.90	9.51	50.22	8.41	48.91	6.73	2.57*	4,117

Note. Groups with different superscripts differed from each other in post-hoc pairwise comparisons, with either Bonferroni or Dunnett T3 corrections, depending on equality of the variances. Note. In reading fluency there were larger *ns*: in Control No-dyslexia *n*=70 in At-risk No -dyslexia *n*= 56, in Late-emerging *n*=15, for the resolving *n*=11, and for the Persistent *n*=18. For the parental cognitive measures *ns* were: in Control No-dyslexia *n*=37 in At-risk No -dyslexia *n*= 43, in Late-emerging *n*=10, for the resolving *n*=10, and for the Persistent *n*=11.

* $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$

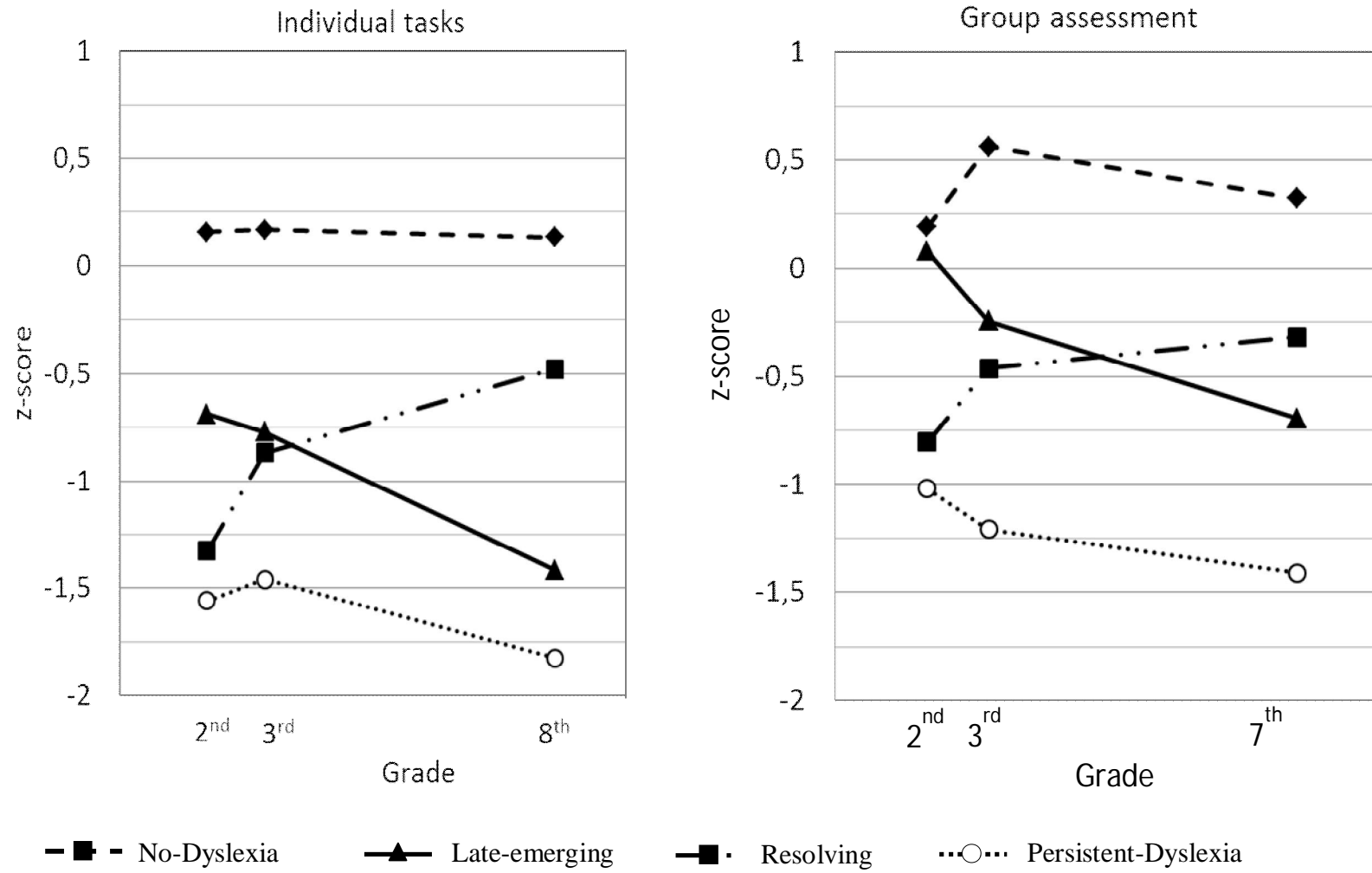


Figure 1. Development of Reading Speed of the RD Groups in Individually Administered and Group Tasks.

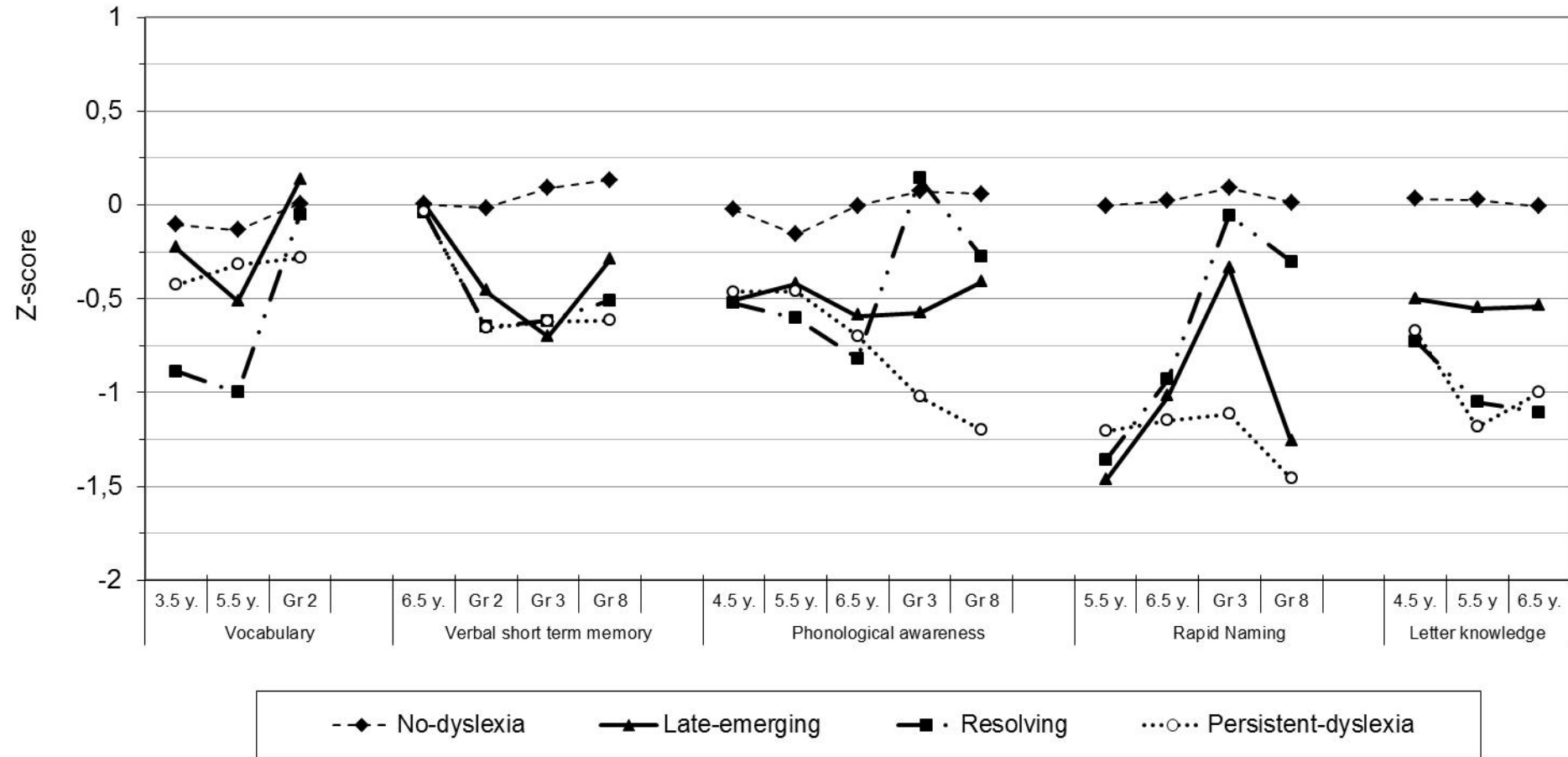


Figure 2. Development of Language and Cognitive Skills in the RD Groups

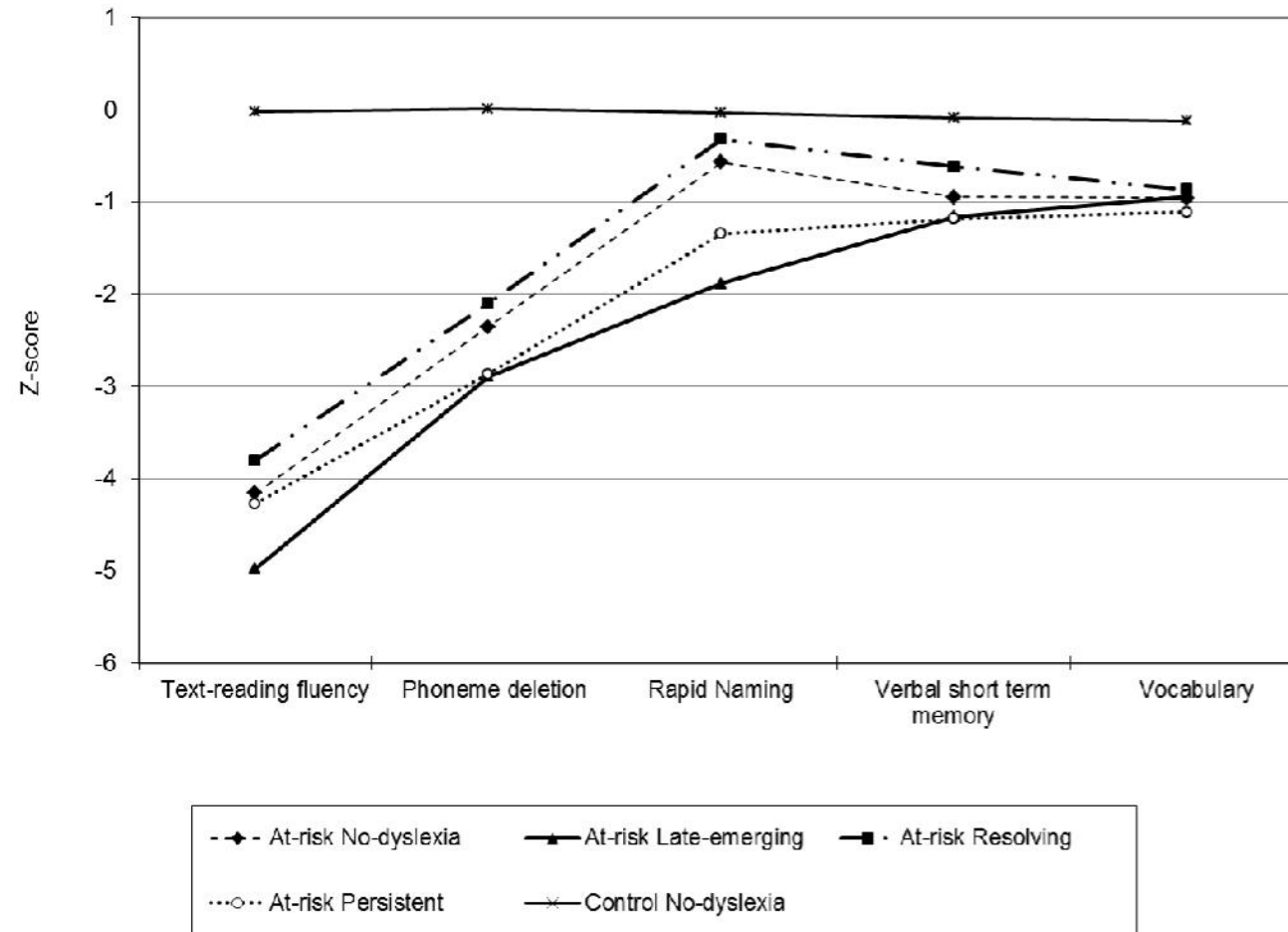


Figure 3. Parental Skills by Risk group and RD Group.