Developmental Science

Developmental Science 2017; 20: e12471 DOI: 10.1111/desc.12471

PAPER

Longitudinal stability of pre-reading skill profiles of kindergarten children: implications for early screening and theories of reading

Ola Ozernov-Palchik,[1,3,5] Elizabeth S. Norton,[2,3,4] Georgios Sideridis,[5,6] Sara D. Beach,[3,4,6] Maryanne Wolf,[1] John D.E. Gabrieli[3,4] and Nadine Gaab[5,6,7]

1. *Eliot-Pearson Department of Child Study and Human Development, Tufts University, USA*
2. *Department of Communication Sciences and Disorders, Northwestern University, USA*
3. *Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, USA*
4. *McGovern Institute for Brain Research, Massachusetts Institute of Technology, USA*
5. *Laboratories of Cognitive Neuroscience, Division of Developmental Medicine, Department of Medicine, Boston Children's Hospital, USA*
6. *Harvard Medical School, USA*
7. *Harvard Graduate School of Education, USA*

Abstract

*Research suggests that early identification of developmental dyslexia is important for mitigating the negative effects of dyslexia, including reduced educational attainment and increased socioemotional difficulties. The strongest pre-literacy predictors of dyslexia are rapid automatized naming (RAN), phonological awareness (PA), letter knowledge, and verbal short-term memory. The relationship among these constructs has been debated, and several theories have emerged to explain the unique role of each in reading ability/disability. Furthermore, the stability of identification of risk based on these measures varies widely across studies, due in part to the different cut-offs employed to designate risk. We applied a latent profile analysis technique with a diverse sample of 1215 kindergarten and pre-kindergarten students from 20 schools, to investigate whether PA, RAN, letter knowledge, and verbal short-term memory measures differentiated between homogenous profiles of performance on these measures. Six profiles of performance emerged from the data: average performers, below average performers, high performers, PA risk, RAN risk, and double-deficit risk (both PA and RAN). A latent class regression model was employed to investigate the longitudinal stability of these groups in a representative subset of children (*n *= 95) nearly two years later, at the end of 1st grade. Profile membership in the spring semester of pre-kindergarten or fall semester of kindergarten was significantly predictive of later reading performance, with the specific patterns of performance on the different constructs remaining stable across the years. There was a higher frequency of PA and RAN deficits in children from lower socioeconomic status (SES) backgrounds. There was no evidence for the IQ–achievement discrepancy criterion traditionally used to diagnose dyslexia. Our results support the feasibility of early identification of dyslexia risk and point to the heterogeneity of risk profiles. These findings carry important implications for improving outcomes for children with dyslexia, based on more targeted interventions.*

Research highlights

- A latent profile analysis revealed heterogeneous profiles of performance on measures of early literacy in 1215 kindergarten/pre-kindergarten students.
- Patterns of performance of the different profiles were in accordance with current theoretical views of dyslexia.
- A 100% stability in profile membership across two years was observed.
- No evidence of a dissociation between general cognitive ability and literacy performance was detected.
- Single-deficit risk profiles, but not the double-deficit profile, were over-represented in the low-SES schools.

Address for correspondence: Ola Ozernov-Palchik, Boston Children's Hospital, Labs of Cognitive Neuroscience, BCH3178, 1 Autumn Street, AU615 Boston, MA 02215, USA; e-mail: ola.ozernov_palchik@tufts.edu

Introduction

Developmental dyslexia (henceforth, dyslexia) affects 5–17% of children, with the prevalence rates varying widely across studies depending on the exact definition and measures used for diagnosis (Elliott & Grigorenko, 2014). Dyslexia is a neurological condition characterized by difficulties with accurate and/or fluent word recognition, poor spelling, and poor decoding abilities (Lyon, Shaywitz & Shaywitz, 2003). Dyslexia is also often associated with impediments in a range of perceptual and cognitive processes important for reading, such as verbal short-term memory, rapid naming, and phonological awareness, as well as differences in the brain regions supporting these processes (Norton, Beach & Gabrieli, 2015; Pennington, Santerre-Lemmon, Rosenberg, MacDonald, Boada *et al.*, 2012). Due to the variability of dyslexia symptoms reported in the literature, forming a cohesive definition of dyslexia has been challenging, and instead a multi-deficit conceptualization of dyslexia is becoming increasingly accepted (Pennington *et al.*, 2012). Traditionally, there has been an emphasis on the independence of dyslexia from other causes that could explain reading failure (i.e., low intelligence, socioeconomic disadvantage, inadequate schooling, or physical disability) (Lyon, 1995). Due to the complex interactions among environmental, cognitive, and neurological factors, however, the merits of such an approach are under considerable debate (Elliott & Grigorenko, 2014).

Similar to the complexity of dyslexia's definition, and potentially because of it, dyslexia remediation efforts have been challenging, with modest effect sizes for interventions ranging from 0.07 to 0.56, according to a meta-analysis (Wanzek & Vaughn, 2007; Wanzek, Vaughn, Scammacca, Metz, Murray *et al.*, 2013). Converging research points to the importance of early and individualized interventions for at-risk students for improving the effectiveness of remediation (Denton & Hocker, 2006; Flynn, Zheng & Swanson, 2012; Morris, Lovett, Wolf, Sevcik, Steinbach *et al.*, 2012; Shaywitz, Morris & Shaywitz, 2008; Swanson & O'Connor, 2009; Torgesen, 2000; Vellutino, Scanlon, Sipay, Small, Pratt *et al.*, 1996; Wanzek & Vaughn, 2007). Therefore, an important question is whether distinct patterns of pre-reading performance could serve as reliable predictors of particular profiles of dyslexia. The current study, for the first time, implements latent profile analysis (LPA) methods to investigate the heterogeneity of dyslexia risk profiles in pre-reading and early-reading children longitudinally. LPA is a mixture-modeling technique that aims to classify individuals into distinct groups based on individual response patterns.

In order to characterize the heterogeneity and prevalence of latent early literacy profiles as well as their longitudinal stability and distribution across school SES levels, a large sample of kindergarten and pre-kindergarten children from 20 diverse schools was evaluated on measures of early literacy and cognition. LPA was implemented to reveal homogenous profiles of performance and to examine these profiles in relation to reading status (readers or pre-readers) and school SES. Latent class membership was then used to predict end-of-1st-grade reading abilities of a subsample of children.

Early identification of dyslexia risk

The cascading effects of early reading ability have been well documented: children who are early readers receive more print exposure and develop superior automaticity, comprehension skills, vocabulary, and cross-domain knowledge (Mol & Bus, 2011; Stanovich, 1996). In contrast, children who lag behind in their early reading abilities receive fewer opportunities to enhance their vocabulary or to develop reading comprehension strategies. In addition, these children tend to acquire negative attitudes about reading (Oka & Paris, 1987), and often remain poor readers throughout their school years and beyond, never achieving fluent reading (Ferrer, Shaywitz, Holahan, Marchione, Michaels *et al.*, 2015; Lyon *et al.*, 2003). Thus, an important aim of reading studies is to determine which pre-reading measures predict dyslexia in order to offer the potential to effectively intervene and prevent reading failure.

Several pre-reading measures, when administered in kindergarten, are predictors of later reading abilities (for a review see Ozernov-Palchik & Gaab, 2016). These measures include letter name and letter sound knowledge (LSK), phonological awareness (PA), verbal short-term memory (VSTM), and rapid automatized naming (RAN) (Catts, Fey, Zhang & Tomblin, 2001; Pennington & Lefly, 2001; Scarborough, 1989; Schatschneider, Fletcher, Francis, Carlson & Foorman, 2004; Wolf, Bally & Morris, 1986). PA is the meta-understanding of the sound units of oral language, measured by the ability to manipulate linguistic sounds independent of meaning (Stahl & Murray, 1994). Short-term memory is a separate, but related, construct that measures the capacity to maintain and process information (e.g., digits, pseudowords) for a short period of time (Siegel & Linder, 1984; Stanovich, Cunningham & Feeman, 1984). VSTM, a short-term memory for linguistic (verbal) material (e.g., a string of letters), is sometimes subsumed under PA, since both involve phonological processing, but there is evidence that it represents a distinct construct and accounts for unique variance in reading (Mann & Liberman, 1984; Scarborough, 1998). RAN is the ability to rapidly retrieve the names of visually presented, familiar items in a serial array (e.g., objects, colors, numbers, or letters, or a combination of these in rapid alternating stimulus formats) (Denckla & Rudel, 1976; Norton & Wolf, 2012).

Although these measures demonstrate a strong association with later reading performance, studies that used kindergarten performance on these measures to ascertain risk for dyslexia showed limited success in predicting which children truly develop dyslexia, with false positives ranging from 20% to 60% (Jenkins & O'Connor, 2002; Torgesen, 2002) and false negatives from 10% to 50% (Catts, 1991; Scarborough, 1998; Torgesen, 2002). These findings prompted suggestions of delaying identification until at least 1st grade, when language-based and literacy-based activities at home are less influential and measures can be more reading-specific (Fletcher, Foorman, Boudousquie, Barnes, Schatschneider *et al.*, 2002). Demonstrating stability in risk-status classification between kindergarten and 1st grade, however, has important implications for early diagnosis of dyslexia risk and, subsequently, early remediation.

Theories of dyslexia and implications for diagnosis and treatment

In the double-deficit view of dyslexia, deficits in PA and RAN represent distinct deficits across different languages; further, the combination of both deficits in some individuals can be additive, creating reading impairment that is more severe than it is in individuals with single deficits (Compton, Defries & Olson, 2001; Kirby, Parrila & Pfeiffer, 2003; O'Brien, Wolf & Lovett, 2012; Papadopoulos, Georgiou & Kendeou, 2009; Wimmer, Mayringer & Landerl, 2000; Wolf & Bowers, 1999). Others contend, however, that rather than representing distinct dyslexia subtypes, both PA and RAN deficits represent the failure to fluently access and retrieve phonological information (Lervåg & Hulme, 2009; Ramus, 2003).

One of the main challenges to the evidence for the double-deficit hypothesis is the inconsistency in criteria applied to designate dyslexia diagnosis across studies (Vukovic & Siegel, 2006). The manner in which deficit groups are defined can strongly influence the results, and thus the understanding of how these deficits relate to reading development. For example, some studies define dyslexia based on an IQ–achievement discrepancy model that designates dyslexia as a low performance on reading assessments relative to performance on tests of general intelligence (e.g., IQ), while others do not. A similar issue concerns the use of arbitrary cut-off criteria to designate risk. For example, across studies, the threshold used to define risk spans the range of the 10th to 25th percentile, or 1 to 2 standard deviations below the standardized mean performance on reading tests. Due to the lack of consensus on the definition of risk, the cut-off method may impose an artificial structure onto data and bias the interpretation of results (Catts, Compton, Tomblin & Bridges, 2012; Fletcher *et al.*, 2002; Francis, Fletcher, Stuebing, Lyon, Shaywitz *et al.*, 2005; Waesche, Schatschneider, Maner, Ahmed & Wagner, 2011). Another consideration is the moderate correlation between PA and RAN that has been shown to impede the methodological validity of classifying children into predetermined discrete PA and RAN deficit subgroups (Compton *et al.*, 2001; Schatschneider, Carlson, Francis, Foorman & Fletcher, 2002).

As an alternative to predefining risk-group membership, several studies used latent analysis methods to ascertain reading profiles or dimensions within large samples of already-reading children. In one study, LPA was used to characterize a large sample of 9-year-old Swedish children on reading performance measures (i.e., reading of continuous texts, reading of document texts, word reading, and reading speed) (Wolff, 2010). Eight stable profiles of readers emerged: (1) high performance; (2) average performance; (3) poor document (e.g., tables, graphs) reading; (4) average decoding, average fluency, poor comprehension (hyperlexic); (5) low decoding, poor fluency, low comprehension (garden-variety poor readers) (6) low decoding, low fluency, poor comprehension (garden-variety poor readers); (7) low-average decoding, low-average fluency, low comprehension; and (8) low decoding, poor fluency, average comprehension (dyslexic).

In a longitudinal study, latent class modeling was also used to identify distinct subtypes of reading development in a large sample of children tested two times per year in the 1st and 2nd grades (Torppa, Tolvanen, Poikkeus, Eklund, Lerkkanen *et al.*, 2007). Several groups of readers emerged based on children's performance on single word identification, reading fluency, and reading comprehension measures: (1) poor readers, (2) slow decoders, (3) poor comprehenders, (4) average readers, and (5) good readers. These studies support the use of data-driven analysis methods for identifying homogenous profiles of reading and suggest that the heterogeneity of reading development is present early in schooling.

In another approach, a taxometric method was applied to identify the latent structure of reading performance in a large sample of 6–8-year-old children who were identified as dyslexic based on IQ–achievement discrepancy or simply low reading achievement (O'Brien *et al.*, 2012). This method allows for revealing latent categorical traits, 'taxons', rather than dimensional classes of the condition of interest in the data. Results from the analysis confirmed the double-deficit view of dyslexia.

However, these results depended on how dyslexia was defined, as the association between RAN and reading emerged in the IQ–achievement discrepancy group, but not in the low reading achievement group. These findings demonstrate the challenge of interpreting results based on pre-established definitions of dyslexia.

The above studies using group classification methods investigated older children who were already reading. From a theoretical perspective, in order to argue that a particular subtype is a core deficit, it is important to demonstrate that the deficit is present prior to reading instruction and is not an artifact of differential influences of reading development, reading instruction, or a phonological awareness deficit (Goswami, 2015). From an applied perspective, the application of these studies to early identification is, therefore, limited.

Longitudinal stability of risk classification

Another important aspect of group classification is the stability of reading profiles across time. Despite the importance of validating classification methods longitudinally, few studies have investigated the stability of specific risk subtypes over time. Instead, most longitudinal investigations have focused on the long-term associations of the early literacy components (i.e., testing whether a particular measure at time 1 correlates with a reading outcome at time 2 (e.g., Scarborough, 1998), or on retrospective investigations of individuals with an existing dyslexia diagnosis (i.e., evaluating performance at time 1 based on outcomes at time 2, e.g., Catts & Weismer, 2006).

The longitudinal stability of PA, RAN, and double-deficit (DD) risk classifications has been investigated in only two studies of pre-readers to date (Spector, 2005; Steacy, Kirby, Parrila & Compton, 2014). Neither study used data-driven methods, but instead applied a predefined cut-off to determine risk. In one study, pre-reading 1st-grade students were classified (using 1 *SD* below mean criterion) as typical, PA deficit, RAN deficit, or DD. These groups exhibited low group membership stability from the beginning to the end of 1st grade (less than 50% accuracy) (Spector, 2005). In another study, kindergarten students were characterized into the same groups using a different cut-off criterion (25th percentile) and were followed longitudinally until the fall of 2nd grade (Steacy *et al.*, 2014). Group membership in this study was highly stable (over 70% accuracy), which might reflect the larger range of scores below the cut-off. Since these two studies applied different criteria to establish risk, the inconsistency in findings could be due to the issues of pre-classifying students into risk groups using cut-off scores. Thus, a data-driven longitudinal approach is necessary to ascertain the stability and utility of a multi-deficit model for dyslexia risk identification.

Cognitive and environmental influences

The complexity of risk identification is underscored by the multiple cognitive and environmental components that interact with reading ability and disability. Two of these components that have been strongly linked to reading performance are intelligence (IQ) and socioeconomic status (SES). Historically, dyslexia has been diagnosed based on a reading achievement and IQ discrepancy model. Several studies have shown that the core mechanisms of dyslexia are consistent regardless of IQ (Siegel, 1989, 1992; Stanovich, 2005; Tanaka, Black, Hulme, Stanley, Kesler *et al.*, 2011), although other studies have demonstrated different patterns of reading profiles based on IQ (Ferrer, Shaywitz, Holahan, Marchione & Shaywitz, 2010; Morris, Stuebing, Fletcher, Shaywitz, Lyon *et al.*, 1998; O'Brien *et al.*, 2012). The relationship between SES and reading achievement is complex, as SES indexes a broad range of environmental factors. Nevertheless, school-level SES (commonly measured by the percentage of children eligible for free/reduced lunch within each school) (Caldas & Bankston, 1997) correlated strongly with reading performance (e.g., Scarborough, 1998; Sirin, 2005). Therefore, both school-SES and IQ are important to consider when investigating reading development.

Current study

The current study aimed to examine: (1) the heterogeneity and prevalence of latent early literacy profiles among kindergarten students; (2) the stability of latent class membership across two time points (i.e., the beginning of kindergarten and the end of 1st grade); and (3) the latent profile distribution across school SES levels. In using LPA and latent class regression methods, we are controlling for some of the major issues raised in the research to date on the presence of single or multiple core deficits in children with dyslexia.

Methods

Participants and data collection procedures

Participating children were recruited from 20 schools in New England. Schools varied in their urbanicity and socioeconomic status and included public district (30%), public charter (20%), private (10%), and Catholic (40%) schools. Schools were classified into three SES groups based on the percentage of students receiving free or reduced lunch, as reported by the school: high SES (0–5% of students, 8 schools), middle SES (12–30%, 6 schools), and low SES (32–79%, 6 schools). Permission and informed consent letters were sent to the parents of kindergarten and pre-kindergarten children to obtain parental permission for their children to participate. Children whose parents provided written informed consent and who provided verbal assent completed a 30–45-minute assessment battery. Assessments were administered by trained research assistants and speech-language pathology students on a 1:1 basis. In total, 1433 English-speaking children were tested at the end of pre-kindergarten or beginning of kindergarten (Year 1). Testing was completed over three years, and therefore the final sample included three cohorts of students. Only children with valid and complete data were included in the current analysis. The final sample included 1215 participants (48% males) with diverse racial (69% Caucasian, 24% African-American/Black, 6% Asian, 1% other) and ethnic (12% Hispanic/Latin) backgrounds. A subset of these children (n = 95, 49% male; 79% Caucasian, 20% African-American/Black, 1% Asian; 5% Hispanic/Latin) was followed longitudinally as part of a larger neuroimaging study and was assessed again at the end of 1st grade (Year 2). Children were recruited to maintain a subsample composition representative of the larger sample with regard to gender, age, ethnicity/race, school type, and behavioral scores. Children with kindergarten IQ scores below 80 and/or who did not speak fluent English, and/or who were born pre-term were excluded from the longitudinal analysis.

Table 1 *Screening and follow-up sample characteristics*

	Mean	SD	Range
Year 1			
Age (months)	65.7	4.2	57–78
KBIT-2 Matrices Standard Score (IQ)	98.9	10.5	80–154
YARC Letter-Sound Knowledge Standard Score (LSK)	99.6	14.8	67–138
CTOPP Elision & Blending Mean Standard Score (PA)	9.9	2	5–18
CTOPP Non-word Repetition Standard Score (VSTM)	8.6	2.6	5–18
RAN Objects & Colors Mean Standard Score (RAN)	97.7	14.7	54–144
WRMT Word ID Raw Score	5.1	11.4	0–71
Year 2			
Age (months)	86.7	4.2	79–100
WISC-IV Processing Speed Standard Score (PS)	10.4	2.7	4–18
TOWRE-2 Sight Word Efficiency Standard Score (SWE)	106.2	14.3	71–138
TOWRE-2 Phonemic Decoding Efficiency Standard Score (PDE)	101.4	14	62–134
CTOPP Elision & Blending Standard Score (PA)	11.7	2.7	7–18
CTOPP Non-word Repetition Standard Score (WM)	9.1	2.2	4–16
RAN Composite (Objects and Colors) Standard Score (RAN)	106.8	15.1	64–145
WRMT-3 Word ID Standard Score	108.3	146	75–145
WRMT-3 Word Attack Standard Score	105.6	13.6	75–135
TWS-5 (SPELL)	106	14.7	71–138
GORT-5 Fluency Standard Score	10.1	2.8	4–17
GORT-5 Comprehension Standard Score	10.1	2.2	6–17

Measures

Group performance on the age-standardized scores of the measures below is summarized in Table 1.

Classification variables (the pre-kindergarten and kindergarten Year 1 screening battery)

Phonological awareness (PA) and verbal short-term memory (VSTM). Three subtests from the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen & Rashotte, 1999) were administered, (1) Elision: the child repeats a word after removing a given syllable or sound; (2) Blending Words: the child blends sounds together to make a real word; (3) Non-word Repetition (NWR): the child repeats a nonsense word. The mean of the Elision and Blending scores was used to calculate the PA composite score.

Rapid automatized naming (RAN). The Colors and Objects subtests of the Rapid Automatized Naming/Rapid Alternating Stimulus (RAN/RAS) tests (Wolf & Denckla, 2005) were administered. The child names an array of familiar items (colors or objects) on the page as quickly and accurately as possible. The raw score is the time to name all items.

Word ID. The Woodcock Reading Mastery Tests – Revised/Normative Update (WRMT-R/NU; Woodcock, 1998) was administered to some children (65%) and the Woodcock Reading Mastery Tests, Third Edition (WRMT-III; Woodcock, 2011) was administered to other children (35%) based on their cohort of participation. For both tests, the Word ID subtest assesses single word reading skills. The child reads aloud single words of increasing difficulty. Scores from the two editions were used interchangeably in the analysis as items early in the test were similar. Children were considered readers based on a raw score of 3 or higher, and non-readers based on a raw score lower than 3. This criterion was chosen based on the median score of 3 for the sample.

Letter sound knowledge (LSK). The Letter Sound Knowledge subtest from the York Assessment of Reading for Comprehension (YARC; Stothard, Hulme, Clarke, Barmby & Snowling, 2010) assesses knowledge of letter sounds. The scores were normed based on the sample distribution in the current study.

Non-verbal IQ (IQ). The Kaufman Brief Intelligence Test, Second Edition (KBIT-2; Kaufman & Kaufman, 2004) Matrices subtest assesses nonverbal matrix reasoning skills, specifically, the understanding of relations between either concrete stimuli (pictures of objects) or abstract stimuli (e.g., designs or symbols).

1st grade (Year 2) assessment included the same measures as in Year 1 and additional measures

Phonemic decoding (Word Attack). The Word Attack subtest of the WRMT-III was administered to measure the ability to apply phonic and structural analysis skills to pronounce unfamiliar words (Woodcock, 2011). The child reads non-words of increasing complexity.

Sight word efficiency (SWE) and phonemic decoding efficiency (PDE). Test of Word Reading Efficiency (TOWRE-2; Torgesen, Wagner & Rashotte, 1999) Sight Word Efficiency and Phonemic Decoding Efficiency subtests were administered to measure word reading accuracy and fluency. A child reads real words or non-words as accurately and as quickly as possible within 45 seconds.

Reading comprehension and fluency. For the Gray Oral Reading Test–5th Edition (GORT-5; Wiederholt & Bryant, 2012), the child reads several stories aloud and then answers questions based on these passages. Scores are determined for rate, comprehension (number of correct comprehension responses), reading accuracy (number of oral reading errors only for the oral reading paragraph), and reading fluency (combination of the rate and accuracy score).

Processing speed (PS). In the Wechsler Intelligence Scale for Children IV-Coding subtest (WISC-IV; Wechsler, 2003), a child copies symbols that are matched with simple shapes within a two-minute time limit.

Spelling (Spell). The Test of Written Spelling (TWS-4; Larsen, Hammill & Moats, 1999) assesses expressive spelling skills. The experimenter reads a word and the child is asked to write it on paper.

Latent Profile Analysis

A latent profile analysis (LPA) approach was employed to identify homogenous subgroups (i.e., profiles) of children based on reading-related variables. Unlike variable-centered approaches (e.g., exploratory factor analysis) that seek to identify correlations between variables of interest, LPA is a person-centered approach that groups individuals by the probability of their response patterns on each of the latent profile indicators. The interpretation of each profile is derived from those probabilities. Specifically, the model-fitting process begins with a one-profile model (i.e., a model in which all readers are hypothesized to demonstrate a single, homogeneous profile) to which additional profiles are added one at a time. Statistical tests are conducted at each step to determine whether the additional profile significantly improves the goodness of fit of the model. Simulation studies in the statistical literature have found that these tests are robust and specific in determining when latent profiles can and cannot be differentiated in the population (e.g., Enders & Tofighi, 2007; Lo, Mendell & Rubin, 2001). Several criteria are employed for testing optimal fit. These include: (a) magnitude of R^2 values; (b) correct classifications versus misclassifications, e.g., in the longitudinal prediction between original class cases and their representation in the predicted classes; (c) significance of predictors; (d) significant reduction in the likelihood ratio test L^2 when comparing nested models; and (e) acceptable entropy values. For continuous indicators such as those involved in the present study, each latent profile was assumed to have its own mean and variance estimates as shown below:

$$f(y|\theta) = \Sigma_{T=1}^{T} \pi_{\tau}^{X} f(y \mid \mu_{\tau}, \Sigma_{\tau}) \tag{1}$$

The distribution of a dependent variable y is a function of a set of unknown parameters θ. On the right side of the equation π defines the probability of person x belonging in latent profile τ. Each latent profile has its own mean (μ_τ) and variance and covariance estimates of the latent profiles (Στ).

In the present study 1–7 profile models were fit to the data and a superior model fit was judged as a function of differences in the likelihood ratio between nested models, using the unbiased bootstrap distribution (Magidson & Vermunt, 2002). Furthermore, parsimony was taken into consideration by selecting the model with the smallest

AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion) values, due to the fact that the likelihood ratio (LR) test will likely be influenced by the large sample size (Akaike, 1974). The level of significance was set to 5% (Tofighi & Enders, 2008).

Sample size estimation in latent profile models involves comparing models and thus relates strongly to the power of the LR test. Recommendations from previous simulation studies have suggested that 500 participants would suffice for highly demanding models (those with minimal between-profile membership differences and small numbers of indicators; Nylund, Asparouhov & Muthén, 2007). Our sample size of 1215 participants was more than adequate for estimating our 6-profile model. As a secondary precaution and because the chi-square test is sensitive to sparse data, it is recommended to bootstrap *p*-value estimates in order to test the improvement in fit between two models using population-based estimates. This approach was followed in the present study using 1000 replicated data sets, thus simulating population parameters based on our large sample. All models were run using Latent Gold 5.0 (Vermunt & Magidson, 2013).

Parameter estimates were presented with effect size metrics, specifically Cohen's *d* statistic (Cohen, 1992). Cohen's *d* is a metric of standard deviations and therefore differences in the latent class membership (Figures 1 and 2) are expressed as standard deviations from the mean of zero (i.e., *z*-scores). As is customary, effect sizes of 0.5 are considered medium size and significant (as would be derived from inferential analyses), effects greater than 0.8 are considered large, and effects between 0 and 0.49 are considered small to medium and non-significant.

Results

Subtypes of early reading profiles

The baseline model estimated a 1-profile solution which formed the basis for subsequent comparisons. When comparing a 2-profile model to the baseline model, the fit of the 2-profile model was superior, but still not acceptable because the classification errors were at the level of significance and the residual values[1] exceeded the 1.0 recommended value (Magidson & Vermunt, 2001, 2002). Subsequently, a 3-profile solution was estimated with the purpose of minimizing those residual covariations and improving model fit (Table 2). Results suggested that the classification errors were still unacceptably high ($p < .05$). Thus, the 3-profile model was discarded in favor of a 4-profile model. When this model was tested using the log-likelihood -2LL (log likelihood) statistic based on the bootstrap distribution and 1000 replications ($-2LL_{Diff} = 357.572$, $p < .001$), it was superior to the 3-profile model. The process was repeated until the superiority of any subsequent model would not be evident using the BIC and/or Consistent AIC index (Tofighi & Enders, 2008). A 5-profile solution was tested and was statistically superior to the 4-profile solution ($-2LL_{Diff} = 278.551$, $p < .001$), with a significant improvement in fit by also employing the BIC and CAIC (Table 2). The superiority of a 6-profile model was tested against the 5-profile model, which was also supported ($-2LL_{Diff} = 158.084$, $p < .001$). However, when moving to a 7-profile model, its superiority was not substantiated. First, the BIC and Consistent AIC values were not improved, suggesting a return to the 6-profile model for reasons of parsimony (i.e. $BIC_{7\text{-class}} = 37677.9844$, $BIC_{6\text{-class}} = 37656.0927$; $CAIC_{7\text{-class}} = 37777.9844$, $CAIC_{6\text{-class}} = 37741.0927$). Thus, the preferred model involved a 6-class solution, which is discussed in detail below.

Profile descriptions based on kindergarten data

The six latent profiles/classes[2] were each defined with ample participants (see *n* for each profile in Table 4). The point estimates of each of the six latent classes across IQ, PA, NWR, RAN and LSK predictors (Table 3) demonstrated that (a) each measure was associated with differential effects (levels) across classes (as shown by the significant Wald tests); and (b) the amount of variance of each predictor explained by the latent class membership was both significant and large as shown by the R^2 values, ranging from 13.3% for IQ to 71.1% for LSK.

The profiles were further characterized in terms of the reading performance (i.e., non-readers versus readers) of their members (Table 4). The order and numbering of the profiles was determined by the number of group members, from largest to smallest.

Profile 1

The 'average performers' group was the largest group and included 378 children (31.1%). This profile was

[1] Reflecting chi-square statistics regarding the conditional independence assumption. They are bivariate correlations of error between pairs of independent variables. Their expected value is 1.0 when no significant correlation is present.

[2] The terms profiles and classes are used interchangeably.

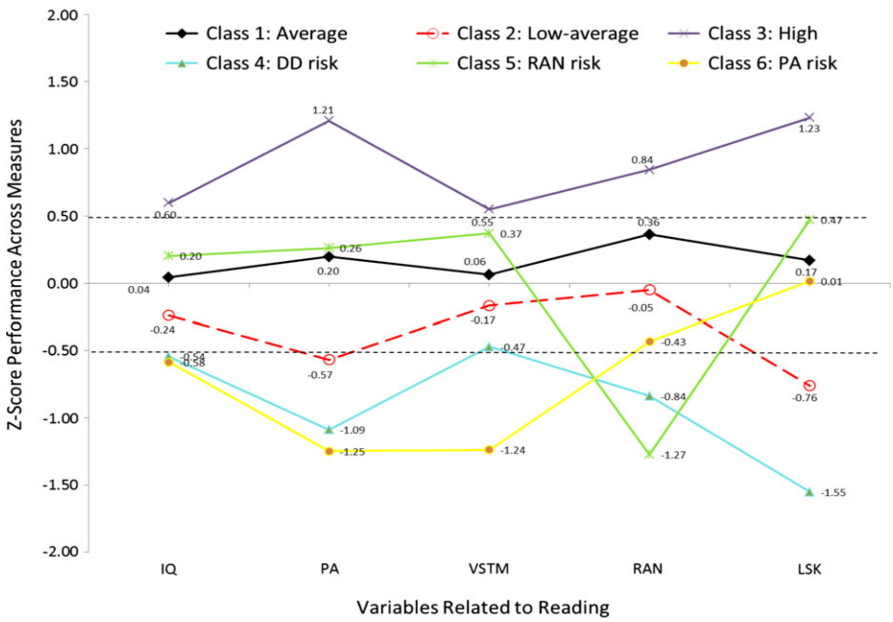

Figure 1 *Latent Profile Analysis model for the identification of reading subgroups: optimal solution. Raw scores were transformed to z-scores on all variables. PA: phonological awareness, VSTM: verbal short-term memory, RAN: rapid automatized naming, LSK: letter sound knowledge.*

associated with performance near the mean score across all measures. Most of the members in this group were non-readers (75.9%), and the group average performance ranged between 0.0 and 0.5 standard deviations[3] from the age-normed test mean across all measures.

Profile 2

The 'low-average performers' group was the second-largest and included 249 children (20.5%). The scores in this group fell slightly below the test mean on all measures except PA and LSK for which they were 0.5 *SD*s and 0.8 *SD*s below the mean, thus representing medium and large effect sizes, respectively. Most children in this group were non-readers (97.2%).

Profile 3

The 'high performers' group included 235 children (19.3%). This group had a similar, but higher, pattern of performance as compared to Profile 1 and was associated with 0.5 to 1 *SD* above average performance across all measures. This high performing group had achievement levels exceeding a medium effect size (i.e., 0.5) across all measures. In all, 89.8% of the members in this group were readers.

Profile 4

The 'double-deficit (DD) risk' group included 147 children (12.1%) and was the poorest-performing class. This class was associated with 0.5 to 1.6 *SD* below the mean performance across all measures. This was the only group in which all members in the group were non-readers (100%).

Profile 5

The 'RAN risk' group included 132 children (10.86%). This class was associated with at-mean or slightly higher than the mean performance on all measures (with effect sizes ranging between 0.2 and 0.5) except RAN. On RAN, the group performed 1.27 *SD* below the mean. Over half (58.3%) of the members of this group were non-readers.

Profile 6

The 'PA risk' group was the smallest group, including 74 children (6.09%). This group performed 0.5 *SD* below the mean on IQ and more than 1 *SD* below the mean on PA and NWR. Their RAN performance was close to a medium effect size (i.e., −0.43) below the mean and their

[3] A standard deviation of 0.5 was selected to represent a medium effect size based on Cohen (1992).

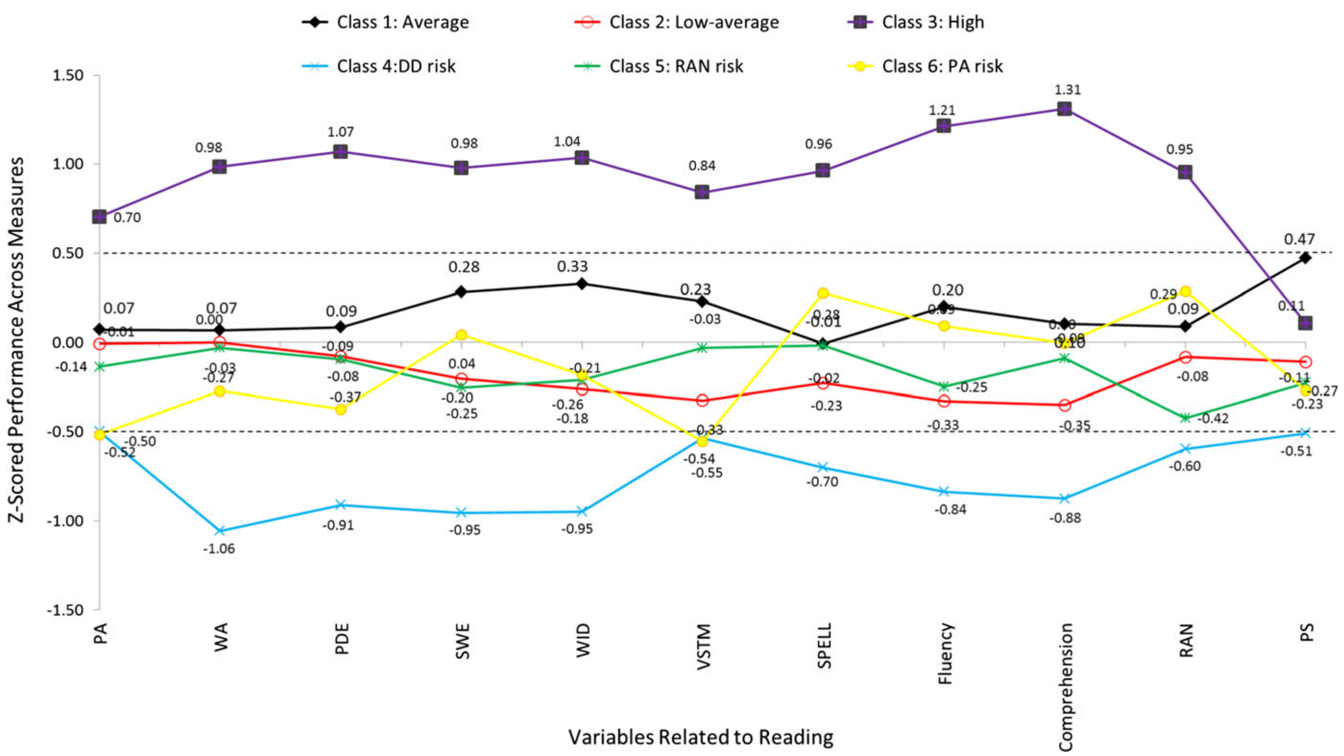

Figure 2 *Latent class regression predicting reading outcomes from the earlier-formed kindergarten latent profiles. The values in the figure are z-scores. The Wald z-statistic suggested that all variables were significant in differentiating between latent classes. That is, for all variables there were differential levels of performance per class. PA: phonological awareness, VSTM: verbal short-term memory, WA: word attack, PDE: phonemic decoding efficiency, SWE: sight word efficiency, WID: word identification, SPELL: spelling, RAN: rapid automatized naming, PS: processing speed.*

LSK performance was average. Most members of this class were non-readers (89.2%).

Differences between classes were evaluated by means of the omnibus Wald test and in case of significance, a series of post-hoc tests. However, because those post-hoc tests were run under excessive power levels, due to the large sample size at Year 1, almost all estimates were significant. In order to avoid inflated statistics, the comparison between classes was based on effect size metrics, specifically, Cohen's d statistic as discussed above.

Longitudinal prediction based on latent class membership

A latent class regression model (Magidson & Vermunt, 2001) was employed to test the hypothesis that profile membership in kindergarten would be predictive of end-of-1st-grade reading performance for the subset of children who participated in the follow-up portion of the study (n = 95). A Monte Carlo simulation was conducted to estimate the power levels of the mixture model using an n-size of 95 participants. The simulation involved a latent profile variable with 11 continuous indicators and six latent classes for which a standardized mean estimate of 0.80 (suggesting a large effect) was tested for significance (through fixing the latent class variances to 1). Results indicated that coverage (i.e., the confidence intervals that contained the true population mean of 0.8) ranged between 80% and 85.8% and power (proportion of correct rejections) at a 5% level of significance ranged between 70% and 74%. Both estimates of coverage and power were acceptable using our proposed sample size of 95 participants.

In addition, a bias analysis was conducted to ensure that the Year 2 cases were allocated among the classes the same way as in the original (kindergarten) sample. This analysis ruled out the possibility that the findings at Year 1 are due to a different composition of the sample at Year 2 compared to that of the kindergarten sample. Specifically, a cross-tabulation analysis using Pearson's chi-square statistic was conducted to evaluate the similarity in percentages. The omnibus Pearson chi-square test was non-significant [$\chi^2(5) = 7.36$, p = .92], indicating a similar representation of cases in the six profiles for the

Table 2 *Model comparison for optimum latent class solution*

Model	LL†	BIC(LL)	AIC(LL)	AIC3(LL)	CAIC(LL)	Npar	Class. Err.
1-Class	−20054.9921	40181.0093	40129.9843	40139.9843	40191.0093	10	0.0000*
2-Class	−19204.064	38585.6904	38458.1279	38483.1279	38610.6904	25	0.0574
3-Class	−18923.2936	38130.6872	37926.5872	37966.5872	38170.6872	40	0.0465*
4-Class	−18744.5077	37879.6529	37599.0155	37654.0155	37934.6529	55	0.0662
5-Class	−18605.2323	37707.6396	37350.4646	37420.4646	37777.6396	70	0.0876
6-Class	***−18526.1901***	***37656.0927***	***37222.3803***	***37307.3803***	***37741.0927***	***85***	***0.0854***
7-Class	−18483.8672	37677.9844	37167.7345	37267.7345	37777.9844	100	0.1011

Note: $^{*}p < .05$; $^{**}p < .01$. Optimum solution is in italics and reflects a 6-class latent variable model. LL = log likelihood; BIC = Bayesian Information Criterion; AIC = Akaike Information Criterion; AIC3 = Corrected AIC with a penalty factor of three; CAIC = Consistent AIC; Npar = Number of estimated parameters; Class. Err. = Classification error. Preferred models should have non-significant amounts of classification errors. †The Bootstrapped Likelihood Ratio Test (BLRT) was employed in order to compare adjacent models using 500 replications. Thus, the 2-class model provided a significant improvement over the 1-class model [−2LL Diff = 992.244, $p < .001$]; similarly the 3-class model was superior to the 2-class model [−2LL Diff = 195.254, $p < .001$] and the 4-class model significantly improved on the 3-class model [−2LL Diff = 539.211, $p < .001$]. The 5-class model improved on the 4-class model [−2LL Diff = 343.624, $p < .001$] and the 6-class model on the 5-class model [−2LL Diff = 48.145, $p < .01$]. The 7-class model was statistically a superior model to the 6-class model but the parsimoniousness indices (BIC and CAIC) suggested that it was over-parameterized in relation to the amount of information it provided.

Table 3 *Point estimates of each of the six latent classes across IQ, PA, NWR, RAN and LSK predictors*

Predictors	Average	Low-average	High	DD risk	RAN risk	PA risk	Wald	R^2
IQ	98.941	96.179	104.112	92.519	100.049	92.796	106.18***	0.133
PA	10.295	8.722	12.316	7.58	10.338	7.249	1457.53***	0.593
NWR	8.758	8.255	10.026	7.382	9.445	5.474	367.83***	0.191
RAN	103.133	97.223	110.054	85.354	79.253	90.798	1037.15***	0.451
LSK	103.523	90.159	118.537	77.714	107.426	100.213	2365.5***	0.711

Note: Values in the table are means in the original score metric, for clarity. $^{***}p < .001$.

Table 4 *Reading performance of the members of each profile*

	Average		Low-average		High		DD risk		RAN risk		PA risk		Total	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Non-readers	287	75.9	242	97.2	24	10.2	147	100.0	77	58.3	66	89.2	843	69.4
Readers	91	24.1	7	2.8	211	89.8	0	0.0	55	41.7	8	10.8	372	30.6
Total*	378	34.0	249	28.7	235	2.8	147	17.4	132	9.1	74	7.8	1215	100.0

*Column percentage totals represent the percentage of non-readers in each class as compared to the total number of non-readers.

longitudinal participants, as compared to the full sample. Specifically, the percentages per class were as follows (Year 1/Year 2): for Class 1, 31%/27%; for Class 2, 20%/23%; for Class 3, 20%/14%; for Class 4, 12%/13%; for Class 5, 10%/18%; and for Class 6, 6%/5%. Consequently, the Year 1 6-class category classification was used as an independent variable and the following Year 2 measures were entered as dependent variables: (a) WISC-IV Coding (WISC PS), (b) TOWRE-2 Sight Word Efficiency (SWE), (c) TOWRE-2 Phonemic Decoding Efficiency (PDE), (d) CTOPP Elision and Blending composite score (PA), (e) CTOPP Non-word Repetition (NWR), (f) WRMT-III Word ID, (g) WRMT-III Word Attack, (h) RAN Objects and Colors composite score, (i) TWS-4 Spelling (SPELL), (j) GORT-5 Fluency, (k) GORT-5 Comprehension. The tested means and significance levels shown in Table 3 suggested that the classes were adequately differentiated based on the classification variables. Table 5 shows the means on each of the Year 1 measures for each class and the Wald statistics, which indicate differences between classes on mean point estimates. Latent class formation was distinct across all

Table 5 *Mean point estimates and significance tests for longitudinal prediction of reading outcomes from latent class formation at kindergarten*

Predictors	Average	Low-average	High	DD risk	RAN risk	PA risk	Wald	R^2
WISC PS	11.779	9.935	10.316	8.916	9.706	9.518	14.264***	0.148
SWE	109.674	103.157	122.754	92.135	101.76	102.895	86.013***	0.327
PDE	101.909	99.729	117.185	87.858	98.763	91.9	56.200***	0.304
PA	11.724	11.724	13.836	10.447	11.099	9.599	18.511***	0.157
NWR	9.425	8.33	10.86	7.895	8.853	7.519	28.588***	0.216
Word ID	112.342	104.168	123.68	94.616	104.081	101.159	80.041***	0.348
Word Attack	105.259	105.368	118.634	90.279	103.318	97.798	48.890***	0.292
RAN	108.27	106.309	121.512	97.122	99.808	109.55	19.849***	0.224
SPELL	104.762	101.785	124.111	92.906	104.451	106.335	34.349***	0.300
Fluency	10.837	9.504	13.885	8.132	9.663	10.125	50.975***	0.383
Comprehension	10.142	9.328	13.078	8.2	9.751	9.549	50.386***	0.413

Note: $^*p < .05$; $^{**}p < .01$; $^{***}p < .001$.

measures, such that no measures were associated with identical point estimates across the latent classes.

A latent class analysis on all Year 2 measures revealed six distinct profiles of performance. The patterns of performance across profiles on variables that overlapped between the two years, as well as on additional variables, closely resembled the pattern of performance on Year 1 measures. A predictive model was developed to test the likelihood that a particular child who belonged to one class in Year 1 will remain in the same class in Year 2. Results (Figure 2 and Table 5) indicate that all of the children were classified in to the same latent class in Year 2 as in Year 1, reflecting 100% stability in class membership.

In terms of general performance on Year 2 measures, children in the high performers profile (Class 3) had the highest means across all measures except on the WISC PS at Year 2, with effect sizes ranging from medium to large. The DD risk profile (Class 4) maintained low performance in 1st grade across measures with medium to large effect sizes. Specifically, this profile performed worse than all other profiles on all measures except PA, whereas the PA risk profile scored the lowest but very close to the DD group's estimates (with effect sizes of −0.52 and −0.50, respectively). The PA risk group (Class 6) maintained low performance on all phonological measures (PA, PDE, and Word Attack) as well as NWR with small and medium effect sizes, and maintained above average performance on RAN and other speeded measures (Fluency and SWE). Furthermore, the average profile (Class 1) demonstrated a slight advantage (small effect size) in performance on the SWE task as compared to the PDE task. This advantage was significantly more pronounced (medium effect size) in the PA risk group and there was no advantage for the RAN risk (Class 5) and the low-average (Class 2) groups. The low-average group demonstrated below average (small to medium effect size) performance on NWR, Word ID, SPELL, and Comprehension and Fluency measures, but not on any of the phonological decoding and awareness measures, for which performance was at average levels. The RAN risk group (Class 5) remained average-performing on all measures except RAN, for which performance was below average (small to medium effect size). The PA risk group outperformed the RAN group on 1st grade speeded reading measures (i.e., SWE, Fluency) with a small effect size. Children in the RAN and DD risk groups were the only ones who demonstrated higher reading comprehension than reading fluency skills.

Cognitive and environmental factors

To evaluate the relationship between SES and class membership, the distribution of profiles across the three school-level SES groups was tested. Chi-square tests revealed a significant ($p < .001$) difference in profile distribution across the three SES groups. Whereas the majority of high-performing and average-performing students (Profiles 1 and 3) belonged to the high (41.8% and 31.1%, respectively) and medium (36.2% and 49.4%, respectively) SES groups (versus 22% and 19.6% in low the SES group), the majority of the PA and RAN risk students (Profiles 5 and 6) belonged to the low-SES group (37.1% and 56.8% versus 32.6% and 17.6 in high SES and 30.3% and 25.7% in middle SES). These results are especially striking considering that the low-SES class had fewer students overall (n = 314 versus n = 457 for high SES and n = 444 for middle SES). The DD and low-average class distribution was proportional to the SES group size (Figure 3). Pearson correlation revealed that performance on the IQ measure was significantly

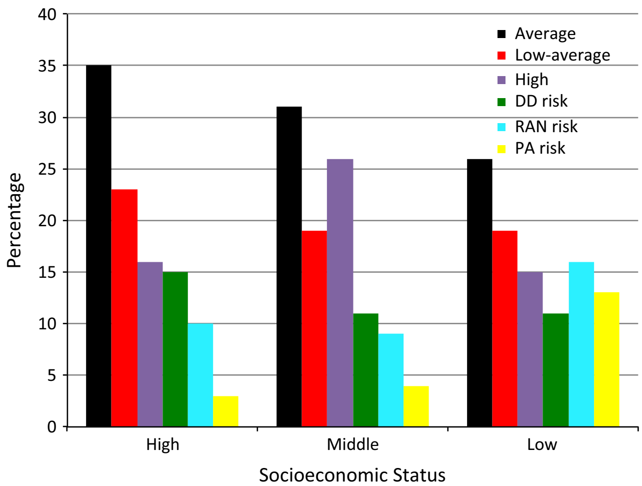

Figure 3 *Latent class distribution across SES groups.*

positively correlated with all Year 1 measures (Pearson r estimates with PA = .344, $p < .001$; with LSK = .140, $p < .001$; with RAN = .226, $p < .001$; and with NWR = .276, $p < .001$).

Discussion

This study was the first to apply latent profile analysis (LPA) and longitudinal regression approaches to characterize the heterogeneous profiles of early reading performance of a large sample of kindergarten and pre-kindergarten students and to evaluate the predictive capacities of these profiles longitudinally in the context of socioeconomic and cognitive factors (i.e., IQ). Six distinct profiles of reading emerged and were characterized as follows: average performers, high performers, low-average performers, RAN risk, PA risk, and double-deficit risk. Importantly, these patterns of performance were in accordance with previous risk classification studies and significantly predicted performance on end-of-1st-grade reading and language measures, revealing a longitudinal stability of class membership of 100%.

Implications for dyslexia risk subtypes

Similar to previous studies that did not use a predetermined cut-off to characterize risk (Boscardin, Muthén, Francis & Baker, 2008; Torppa *et al*., 2007; Wolff, 2010), multiple reading profiles emerged in our sample. Three distinct profiles of deficits that differed in performance level and pattern were identified: PA deficit, RAN deficit, and double-deficit (DD). In terms of general performance on all measures, the DD group performed more poorly than the PA risk group, which in turn had lower scores than the RAN risk group. These results are in line with previous double-deficit studies that found similar relative performance among the PA, RAN, and DD groups (Katzir, Kim, Wolf, Morris & Lovett, 2008; Lovett, Lacerenza, Borden, Frijters, Steinbach *et al*., 2000; Vaessen, Gerretsen & Blomert, 2009; Wolf & Bowers, 1999).

Whereas previous findings of lower PA scores in the DD group as compared to the PA deficit group led some authors to question the validity of the double-deficit distinction (Compton *et al*., 2001; Schatschneider *et al*., 2002), the DD group in our sample had comparable PA scores to the PA risk group. In fact, on the verbal short-term memory measure, the DD group scored significantly higher than the PA group, albeit with a small effect size. This suggests that the reduced performance of the DD group in both years is due to the cascading effects of both phonological and RAN deficits that impair reading acquisition across several levels of processing, e.g., phonological, visual, attentional, and retrieval (Wolf & Bowers, 1999, 2000).

In contrast, the RAN risk group had intact performance on all other kindergarten measures (including PA), further supporting the independence of the RAN construct from PA. RAN is thought to index the automaticity with which cognitive processes important for reading are executed and integrated (Norton & Wolf, 2012). Consequently, RAN has been strongly linked to timed word identification measures and reading fluency. Indeed, the RAN risk group performed below the other profiles (except DD risk) on 1st-grade rate-related skills (i.e., Sight Word Efficiency, Fluency). In addition, the RAN risk group's pattern of low fluency performance as compared to comprehension is in contrast to that of the other groups that demonstrated similar performance on comprehension and fluency. Thus, the current results bolster the specificity of RAN's association to speeded and fluency-related measures.

The low performance of the PA risk group on phonological measures, but not reading or spelling measures, both supports the stability of the PA construct and suggests that the PA deficit on its own is insufficient to cause reading impairment. PA indexes the ability to decode (i.e., sound out) words that are not yet automatic as well as non-words (Stahl & Murray, 1994). Accordingly, the PA risk group had lower phonemic decoding skills (Phonemic Decoding Efficiency, Word Attack) as compared to Sight Word Efficiency, showing a different pattern from the DD and RAN risk groups. The PA risk group's impairment on phonological measures was specific, as they did not show reduced LSK in kindergarten or impaired 1stgrade reading comprehension and fluency performance. This is in line with studies demonstrating a dissociation between phonological deficits and reading performance in the absence of other exacerbating risk factors (Moll, Loff & Snowling, 2013). The small size of the PA risk profile further suggests the rarity of pure phonological deficits early in reading development.

The low-average profile comprised the largest group in the sample and was characterized by below average performance on all kindergarten measures. This group also was characterized by low performance on LSK in kindergarten (as compared to the other measures) and by the non-reading status of the majority of the group. In Year 2, this group demonstrated below-average performance on single word identification measures (Word ID, Sight Word Efficiency), as well as on spelling, reading comprehension, and reading fluency. In both years, the low-average group demonstrated low performance on the verbal short-term memory measure. This unique pattern of poor performance on orthographic measures in kindergarten and typical performance on phonological as compared to orthographic and lexical reading measures in 1st grade, is reminiscent of another conceptualization of dyslexia reported in literature: the surface deficit of dyslexia (Castles & Coltheart, 1993). Surface dyslexia has been characterized by intact phonological abilities and intact regular word reading, but poor exception word reading (Coltheart, Masterson, Byng, Prior & Riddoch, 1983). Exception words are words that have irregular spelling and, therefore, cannot be read by applying phonological grapheme-to-phoneme conversion rules. Instead, these words are read holistically through direct access to the lexical information underlying a specific orthographic pattern. Children with a surface-dyslexia-like deficit have a problem in developing direct visual representations of words and are thus differentially impaired at tasks emphasizing orthographic knowledge (Jiménez, Rodríguez & Ramírez, 2009; Manis, Seidenberg & Doi, 1999; Manis, Seidenberg, Doi, McBride-Chang & Petersen, 1996; Stanovich, Siegel & Gottardo, 1997). Accordingly, the Sight Word Efficiency measure, on which the low-average group showed lower performance as compared to Phonemic Decoding Efficiency, included many irregular words.

There has been mixed evidence for the validity of the surface dyslexia subtype. Some studies with reading-level controls suggested that it represents a developmental delay rather than a distinct deficit (Manis *et al.*, 1996; but see Peterson, Pennington & Olson, 2013; Stanovich *et al.*, 1997). These developmental delays have been attributed to poor home literacy or language environments (Castles, Datta, Gayan & Olson, 1999; Sprenger-Charolles, Siegel, Jiménez & Ziegler, 2011). In addition, studies suggest that phonological and surface dyslexia differ only in the degree of severity of phonological deficits and in cognitive resources available to compensate for these deficits (Snowling, 1998). Future studies will determine whether the initial orthographic deficits demonstrated in the current study for the low-average group will be ameliorated with additional reading instruction or become more pronounced in later grades.

Longitudinal stability of risk classifications

The longitudinal stability of early pre-reading literacy profiles has important implications for dyslexia risk identification and intervention. Our results demonstrated perfect stability in classification from the beginning of kindergarten to the end of 1st grade. Importantly, the patterns of performance on pre-reading measures across the groups correlated with performance on more advanced reading measures in a manner that is consistent with the theoretical expectations of the double-deficit and the surface-phonology deficit approaches. Since children who are poor readers in 1st grade tend to remain poor readers by the end of elementary school (Boscardin *et al.*, 2008; Francis, Shaywitz, Stuebing, Shaywitz & Fletcher, 1996; Juel, 1988; Shaywitz, Fletcher, Holahan, Shneider, Marchione *et al.*, 1999; Torgesen & Burgess, 1998) and on through 12th grade (Ferrer *et al.*, 2015), these findings point to the validity of our kindergarten battery of tests in identifying dyslexia risk and its sensitivity to individual differences in performance.

Indeed, letter knowledge, phonological awareness, verbal short-term memory, and rapid automatized naming have been identified across several studies as the most robust early predictors of reading abilities (Kirby, Desrochers, Roth & Lai, 2008; Ozernov-Palchik & Gaab, 2016; Scarborough, Dobrich & Hager, 1991; Schatschneider *et al.*, 2002; Warmington & Hulme, 2012). Our study demonstrated the stochastic independence among these measures and their robustness in distinguishing between and among various profiles of reading development. Importantly, the differences in performance between typical groups and risk groups on pre-literacy measures extended to differences in actual reading performance on word recognition, fluency, and comprehension measures. These findings suggest that early identification of dyslexia risk is possible and that one-size-fits-all interventions will likely be less effective in accommodating the specific deficits and strengths of the various risk profiles (Allor, Mathes, Jones, Champlin & Cheatham, 2010; Vaughn, Wexler, Leroux, Roberts, Denton *et al.*, 2012).

It is important to note that the theoretical interpretation of our findings could be affected by the selection of measures. The inclusion of other measures in the kindergarten battery could have resulted in different profiles of performance in accordance with other theories of dyslexia (e.g., visual attention). Yet, the selection of measures for the current study was motivated by the robust empirical support for their strong predictive value for reading outcomes across languages, supporting the significance of the current findings. In addition, since the LSK measure was administered in Year 1 only, it was not possible to evaluate the longitudinal stability in performance on this measure. Due to the well-documented limited power of LSK to differentiate between reading abilities beyond kindergarten (due to ceiling effects), the measure was excluded from the Year 2 battery (McBride-Chang, 1999; Wagner & Barker, 1994). Single word measures, however, were administered in Year 2 and are considered a good proxy of early letter knowledge as there is a high concurrent and predictive correlation between the two constructs (Scarborough, 1998; Schatschneider *et al.*, 2004). Indeed, in the current study, children demonstrated similar performance on LSK in Year 1 and Word ID in Year 2. Furthermore, the small size of the longitudinal sample (n = 95) raises the possibility of Type-II error. A Monte Carlo simulation was conducted to test this possibility and showed that such an error was unlikely. By employing a bias analysis, we further demonstrated that the longitudinal sample was representative of the kindergarten sample both in terms of the pattern of distribution across profiles and in demographic characteristics.

Cognitive and environmental factors

Reading development occurs in the context of cognitive and environmental influences. We observed that the frequency of PA and RAN risk was significantly higher in the low-SES schools than in the middle-SES or high-SES schools. This was not the case, however, for DD risk. It is possible that social factors have a higher impact on the single-deficit groups, whereas the double-deficit is influenced more by hereditary factors. Indeed, previous studies reported a higher frequency of a family history of dyslexia in the DD group as compared to other reading profiles (Morris *et al.*, 1998) and studies have demonstrated more severe reading deficits in children with higher genetic liability for dyslexia (van Bergen, van der Leij & de Jong, 2014). Since the majority of schools in this study were charter or private schools, however, students in these schools, even with free/reduced lunch qualification, may not be representative of low-income children who attend non-charter public schools (Lubienski & Lubienski, 2006). For example, in many cases, parents must put forth substantial effort to gain admission and scholarships to a private school or to secure a spot for their child in an oversubscribed charter school. These parents may be more invested in their children's early literacy development. Therefore, different home and school environments may underscore differences in literacy development in children in charter versus public schools. Thus, the current SES results should be interpreted with caution, until future investigations can focus on both school-level and family-level socioeconomic and environmental factors, using a higher proportion of low-SES public schools and a family-level measure of SES.

Our results did not provide support for the IQ–discrepancy model of dyslexia. The PA and DD groups had low-average performance on the non-verbal IQ measure, while the RAN group had average performance. Thus, across the deficit subgroups, the pattern of low language skills despite average IQ did not emerge. In fact, the non-verbal IQ scores were significantly correlated with all Year 1 language measures, indicating a strong coupling between general cognitive abilities and reading. This is in line with previous reports of a strong relationship between cognitive and reading abilities in early grades and the gradual weakening of this relationship across development and into adulthood (Ferrer *et al.*, 2010). Taken together, the current results join an increasing body of evidence against using an IQ-based discrepancy criterion to classify dyslexia risk (Fletcher, Shaywitz, Shankweiler, Katz, Liberman *et al.*, 1994; Pennington, Gilger, Olson & DeFries, 1992; Siegel, 1992; Stanovich & Siegel, 1994; Tanaka *et al.*, 2011). Future investigations should examine how the interaction between general cognitive abilities and reading achievement is mediated by social and hereditary factors in order to best determine particular profiles of reading and dyslexia.

Summary

These findings are novel in applying a data-driven analysis approach to demonstrate the robustness of RAN, PA, VSTM, and LSK tasks administered in early kindergarten in differentiating the discrete subtypes of dyslexia and predicting later reading performance with high accuracy. Current results carry important implications for improved early identification, differentiated remediation, and an evolving understanding of dyslexia. The high stability of group membership supports the feasibility of early identification of risk, prior to reading failure. This is important for optimizing the educational and psychosocial outcomes of children with dyslexia. Performance on the non-verbal IQ measure of the different groups was proportional to the general level of performance across measures, showing no supporting evidence for the IQ–discrepancy model of dyslexia. Finally, the over-representation in low-SES schools of PA and RAN deficit profiles, but not double-deficit or surface-deficit profiles, provides insight both into the environmental factors influencing dyslexia risk, and also possible hereditary factors.

Acknowledgements

This work was supported by grants from the National Institutes of Health–National Institute of Child Health and Human Development (Grant #R01 HD067312 to NG and JDEG, Grant #RO1 HD65762 to NG) and by the Evans Literacy Fellowship to OOP. We thank our research testers, school coordinators and principals, and participating families. Participating schools are listed on our website: http://gablab.mit.edu/index.php/READ-study

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, **19** (6), 716–723.

Allor, J.H., Mathes, P.G., Jones, F.G., Champlin, T.M., & Cheatham, J.P. (2010). Individualized research-based reading instruction for students with intellectual disabilities: success stories. *Teaching Exceptional Children*, **42** (3), 6–12.

Boscardin, C.K., Muthén, B., Francis, D.J., & Baker, E.L. (2008). Early identification of reading difficulties using heterogeneous developmental trajectories. *Journal of Educational Psychology*, **100** (1), 192–208.

Caldas, S.J., & Bankston, C. (1997). Effect of school population socioeconomic status on individual academic achievement. *Journal of Educational Research*, **90** (5), 269–277.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition*, **47** (2), 149–180.

Castles, A., Datta, H., Gayan, J., & Olson, R.K. (1999). Varieties of developmental reading disorder: genetic and environmental influences. *Journal of Experimental Child Psychology*, **72** (2), 73–94.

Catts, H.W. (1991). Early identification of dyslexia: evidence from a follow-up study of speech-language impaired children. *Annals of Dyslexia*, **41** (1), 163–177.

Catts, H.W., Compton, D., Tomblin, J.B., & Bridges, M.S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology*, **104** (1), 166–181.

Catts, H.W., Fey, M.E., Zhang, X., & Tomblin, J.B. (2001). Estimating the risk of future reading difficulties in kindergarten children: a research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools*, **32** (1), 38–50.

Catts, H.W., & Weismer, S.E. (2006). Language deficits in poor comprehenders: a case for the simple view of reading. *Journal of Speech, Language, and Hearing Research*, **49** (2), 278–293.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, **112** (1), 155–159.

Coltheart, M., Masterson, J., Byng, S., Prior, M., & Riddoch, J. (1983). Surface dyslexia. *Quarterly Journal of Experimental Psychology*, **35** (3), 469–495.

Compton, D.L., Defries, J.C., & Olson, R.K. (2001). Are RAN- and phonological awareness-deficits additive in children with reading disabilities? *Dyslexia*, **7** (3), 125–149.

Denckla, M.B., & Rudel, R.G. (1976). Rapid 'automatized' naming (RAN): dyslexia differentiated from other learning disabilities. *Neuropsychologia*, **14** (4), 471–479.

Denton, C., & Hocker, J. (2006). *Responsive reading instruction: Flexible intervention for struggling readers in the early grades.* Longmont, CO: Sopris West Educational Services.

Elliott, J.G., & Grigorenko, E.L. (2014). *The dyslexia debate.* New York: Cambridge University Press.

Enders, C.K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological Methods*, **12** (2), 121–138.

Ferrer, E., Shaywitz, B.A., Holahan, J.M., Marchione, K., & Shaywitz, S.E. (2010). Uncoupling of reading and IQ over time: empirical evidence for a definition of dyslexia. *Psychological Science*, **21** (1), 93–101.

Ferrer, E., Shaywitz, B.A., Holahan, J.M., Marchione, K.E., Michaels, R., et al. (2015). Achievement gap in reading is present as early as first grade and persists through adolescence. *Journal of Pediatrics*, **167** (5), 1121–1125. e1122.

Fletcher, J.M., Foorman, B.R., Boudousquie, A., Barnes, M.A., Schatschneider, C., et al. (2002). Assessment of reading and learning disabilities: a research-based intervention-oriented approach. *Journal of School Psychology*, **40** (1), 27–63.

Fletcher, J.M., Shaywitz, S.E., Shankweiler, D.P., Katz, L., Liberman, I.Y., et al. (1994). Cognitive profiles of reading disability: comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology*, **86** (1), 6–23.

Flynn, L.J., Zheng, X., & Swanson, H.L. (2012). Instructing struggling older readers: a selective meta-analysis of intervention research. *Learning Disabilities Research & Practice*, **27** (1), 21–32.

Francis, D.J., Fletcher, J.M., Stuebing, K.K., Lyon, G.R., Shaywitz, B.A., et al. (2005). Psychometric approaches to the identification of LD: IQ and achievement scores are not sufficient. *Journal of Learning Disabilities*, **38** (2), 98–108.

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** (1), 3–17.

Goswami, U. (2015). Sensory theories of developmental dyslexia: three challenges for research. *Nature Reviews Neuroscience*, **16** (1), 43–54.

Jenkins, J.R., & O'Connor, R.E. (2002). Early identification and intervention for young children with reading/learning disabilities. In R. Bradley, L. Danielson & D.P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 99–149). Mahwah, NJ: Erlbaum.

Jiménez, J.E., Rodríguez, C., & Ramírez, G. (2009). Spanish developmental dyslexia: prevalence, cognitive profile, and home literacy experiences. *Journal of Experimental Child Psychology*, **103** (2), 167–185.

Juel, C. (1988). Learning to read and write: a longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, **80** (4), 437–447.

Katzir, T., Kim, Y.-S., Wolf, M., Morris, R., & Lovett, M.W. (2008). The varieties of pathways to dysfluent reading: comparing subtypes of children with dyslexia at letter, word, and connected text levels of reading. *Journal of Learning Disabilities*, **41** (1), 47–66.

Kaufman, A.S., & Kaufman, N.L. (2004). *Kaufman Brief Intelligence Test*. Wiley Online Library.

Kirby, J.R., Desrochers, A., Roth, L., & Lai, S.S. (2008). Longitudinal predictors of word reading development. *Canadian Psychology/Psychologie canadienne*, **49** (2), 103–110.

Kirby, J.R., Parrila, R.K., & Pfeiffer, S.L. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology*, **95** (3), 453–464.

Larsen, S., Hammill, D., & Moats, L. (1999). *Test of Written Spelling-Revised, IV.* San Antonio, TX: Pearson.

Lervåg, A., & Hulme, C. (2009). Rapid automatized naming (RAN) taps a mechanism that places constraints on the development of early reading fluency. *Psychological Science*, **20** (8), 1040–1048.

Lo, Y., Mendell, N.R., & Rubin, D.B. (2001). Testing the number of components in a normal mixture. *Biometrika*, **88** (3), 767–778.

Lovett, M.W., Lacerenza, L., Borden, S.L., Frijters, J.C., Steinbach, K.A., et al. (2000). Components of effective remediation for developmental reading disabilities: combining phonological and strategy-based instruction to improve outcomes. *Journal of Educational Psychology*, **92** (2), 263–283.

Lubienski, C., & Lubienski, S.T. (2006). *Charter, private, public schools and academic achievement: New evidence from NAEP mathematics data (Vol. 16)*. New York: National Center for the Study of Privatization in Education, Teachers College, Columbia University.

Lyon, G.R. (1995). Toward a definition of dyslexia. *Annals of Dyslexia*, **45** (1), 1–27.

Lyon, G.R., Shaywitz, S.E., & Shaywitz, B.A. (2003). A definition of dyslexia. *Annals of Dyslexia*, **53** (1), 1–14.

McBride-Chang, C. (1999). The ABCs of the ABCs: the development of letter name and letter sound knowledge. *Merrill-Palmer Quarterly*, **45** (2), 7.

McBride-Chang, C., & Manis, F.R. (1996). Structural invariance in the associations of naming speed, phonological awareness, and verbal reasoning in good and poor readers: a test of the double deficit hypothesis. *Reading and Writing*, **8** (4), 323–339.

Magidson, J., & Vermunt, J.K. (2001). Latent Class Factor and Cluster models, bi-plots, and related graphical displays. *Sociological Methodology*, **31** (1), 223–264.

Magidson, J., & Vermunt, J. (2002). Latent class models for clustering: a comparison with K-means. *Canadian Journal of Marketing Research*, **20** (1), 36–43.

Manis, F.R., Seidenberg, M.S., & Doi, L.M. (1999). See Dick RAN: rapid naming and the longitudinal prediction of reading subskills in first and second graders. *Scientific Studies of Reading*, **3** (2), 129–157.

Manis, F.R., Seidenberg, M.S., Doi, L.M., McBride-Chang, C., & Petersen, A. (1996). On the bases of two subtypes of development dyslexia. *Cognition*, **58** (2), 157–195.

Mann, V.A., & Liberman, I.Y. (1984). Phonological awareness and verbal short-term memory. *Journal of Learning Disabilities*, **17** (10), 592–599.

Mol, S.E., & Bus, A.G. (2011). To read or not to read: a meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, **137** (2), 267–296.

Moll, K., Loff, A., & Snowling, M.J. (2013). Cognitive endophenotypes of dyslexia. *Scientific Studies of Reading*, **17** (6), 385–397.

Morris, R., Lovett, M., Wolf, M., Sevcik, R., Steinbach, K., et al. (2012). Multiple component remediation of developmental reading disabilities: a controlled factorial evaluation of the influence of IQ, socioeconomic status, and race on outcomes. *Journal of Learning Disabilities*, **45** (2), 99–127.

Morris, R.D., Stuebing, K.K., Fletcher, J.M., Shaywitz, S.E., Lyon, G.R., et al. (1998). Subtypes of reading disability: variability around a phonological core. *Journal of Educational Psychology*, **90** (3), 347–373.

Norton, E.S., Beach, S.D., & Gabrieli, J.D. (2015). Neurobiology of dyslexia. *Current Opinion in Neurobiology*, **30**, 73–78.

Norton, E.S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: implications for understanding and treatment of reading disabilities. *Annual Review of Psychology*, **63**, 427–452.

Nylund, K.L., Asparouhov, T., & Muthén, B.O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. *Structural Equation Modeling*, **14** (4), 535–569.

O'Brien, B.A., Wolf, M., & Lovett, M.W. (2012). A taxometric investigation of developmental dyslexia subtypes. *Dyslexia*, **18** (1), 16–39.

Oka, E., & Paris, S. (1987). Patterns of motivation and reading skills in underachieving children. In S.J. Ceci (Ed.), *Handbook of cognitive, social, and neuropsychological aspects of learning disabilities* (**Vol. 2**, pp. 115–145). Hillsdae, NJ: Lwarence Erlbaum Associates.

Ozernov-Palchik, O., & Gaab, N. (2016). Tackling the 'dyslexia paradox': reading brain and behavior for early markers of developmental dyslexia. *Wiley Interdisciplinary Reviews: Cognitive Science*, **7** (2), 156–176.

Papadopoulos, T.C., Georgiou, G.K., & Kendeou, P. (2009). Investigating the double-deficit hypothesis in Greek: findings from a longitudinal study. *Journal of Learning Disabilities*, **40**, 319–330.

Pennington, B.F., Gilger, J.W., Olson, R.K., & DeFries, J.C. (1992). The External validity of age-versus IQ-discrepancy definitions of reading disability lessons from a twin study. *Journal of Learning Disabilities*, **25** (9), 562–573.

Pennington, B.F., & Lefly, D.L. (2001). Early reading development in children at family risk for dyslexia. *Child Development*, **72** (3), 816–833.

Pennington, B.F., Santerre-Lemmon, L., Rosenberg, J., MacDonald, B., Boada, R., et al. (2012). Individual prediction of dyslexia by single versus multiple deficit models. *Journal of Abnormal Psychology*, **121** (1), 212–224.

Peterson, R.L., Pennington, B.F., & Olson, R.K. (2013). Subtypes of developmental dyslexia: testing the predictions of the dual-route and connectionist frameworks. *Cognition*, **126** (1), 20–38.

Ramus, F. (2003). Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology*, **13** (2), 212–218.

Scarborough, H.S. (1989). Prediction of reading disabilites from familial and individual differences. *Journal of Educational Psychology*, **81** (1), 101–108.

Scarborough, H.S. (1998). Predicting the future achievement of second graders with reading disabilities: contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia*, **48** (1), 115–136.

Scarborough, H.S., Dobrich, W., & Hager, M. (1991). Preschool literacy experience and later reading achievement. *Journal of Learning Disabilities*, **24** (8), 508–511.

Schatschneider, C., Carlson, C.D., Francis, D.J., Foorman, B.R., & Fletcher, J.M. (2002). Relationship of rapid automatized naming and phonological awareness in early reading development: implications for the double-deficit hypothesis. *Journal of Learning Disabilities*, **35** (3), 245–256.

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** (2), 265–282.

Shaywitz, S.E., Fletcher, J.M., Holahan, J.M., Shneider, A.E., Marchione, K.E., et al. (1999). Persistence of dyslexia: the Connecticut longitudinal study at adolescence. *Pediatrics*, **104** (6), 1351–1359.

Shaywitz, S.E., Morris, R., & Shaywitz, B.A. (2008). The education of dyslexic children from childhood to young adulthood. *Annual Review of Psychology*, **59**, 451–475.

Siegel, L.S. (1989). IQ is irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities*, **22** (8), 469–478.

Siegel, L.S. (1992). An evaluation of the discrepancy definition of dyslexia. *Journal of Learning Disabilities*, **25** (10), 618–629.

Siegel, L.S., & Linder, B.A. (1984). Short-term memory processes in children with reading and arithmetic learning disabilities. *Developmental Psychology*, **20** (2), 200–207.

Sirin, S.R. (2005). Socioeconomic status and academic achievement: a meta-analytic review of research. *Review of Educational Research*, **75** (3), 417–453.

Snowling, M. (1998). Dyslexia as a phonological deficit: evidence and implications. *Child Psychology and Psychiatry Review*, **3** (1), 4–11.

Spector, J.E. (2005). Instability of double-deficit subtypes among at-risk first grade readers. *Reading Psychology*, **26** (3), 285–312.

Sprenger-Charolles, L., Siegel, L.S., Jiménez, J.E., & Ziegler, J.C. (2011). Prevalence and reliability of phonological, surface, and mixed profiles in dyslexia: a review of studies conducted in languages varying in orthographic depth. *Scientific Studies of Reading*, **15** (6), 498–521.

Stahl, S.A., & Murray, B.A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology*, **86** (2), 221–234.

Stanovich, K.E. (1996). Toward a more inclusive definition of dyslexia. *Dyslexia*, **2** (3), 154–166.

Stanovich, K.E. (2005). The future of a mistake: will discrepancy measurement continue to make the learning disabilities field a pseudoscience? *Learning Disability Quarterly*, **28**, 103–106.

Stanovich, K.E., Cunningham, A.E., & Feeman, D.J. (1984). Relation between early reading acquisition and word decoding with and without context: a longitudinal study of first-grade children. *Journal of Educational Psychology*, **76** (4), 668–677.

Stanovich, K.E., & Siegel, L.S. (1994). Phenotypic performance profile of children with reading disabilities: a regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology*, **86** (1), 24–53.

Stanovich, K.E., Siegel, L.S., & Gottardo, A. (1997). Converging evidence for phonological and surface subtypes of reading disability. *Journal of Educational Psychology*, **89** (1), 114–127.

Steacy, L.M., Kirby, J.R., Parrila, R., & Compton, D.L. (2014). Classification of double deficit groups across time: an analysis of group stability from kindergarten to second grade. *Scientific Studies of Reading*, **18** (4), 255–273.

Stothard, S., Hulme, C., Clarke, P., Barmby, P., & Snowling, M. (2010). *YARC York Assessment of Reading for Comprehension (Secondary)*. London: G.L. Assessment.

Swanson, H.L., & O'Connor, R. (2009). The role of working memory and fluency practice on reading comprehension of students who are dysfluent readers. *Journal of Learning Disabilities*, **42** (6), 548–575.

Tanaka, H., Black, J.M., Hulme, C., Stanley, L.M., Kesler, S.R., et al. (2011). The brain basis of the phonological deficit in dyslexia is independent of IQ. *Psychological Science*, **22** (11), 1442–1451.

Tofighi, D., & Enders, C.K. (2008). Identifying the correct number of classes in growth mixture models. In G.R. Hancock & K.M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 317–341). Greenwich, CT: Information Age.

Torgesen, J.K. (2000). Individual differneces in response to early interventions in reading: the lingering problem of treatment resisters. *Learning Disabilities Research & Practice*, **15** (1), 55–64.

Torgesen, J.K. (2002). The prevention of reading difficulties. *Journal of School Psychology*, **40** (1), 7–26.

Torgesen, J.K., & Burgess, S.R. (1998). Consistency of reading-related phonological processes throughout early childhood:

evidence from longitudinal-correlational and instructional studies. In J.L. Matsala & L.C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 161–188). Mahwah, NJ: Erlbaum.

Torgesen, J.K., Wagner, R.K., & Rashotte, C.A. (1999). *Test of word reading efficiency.* Austin, TX: PRO-ED.

Torppa, M., Tolvanen, A., Poikkeus, A.-M., Eklund, K., Lerkkanen, M.-K., et al. (2007). Reading development subtypes and their early characteristics. *Annals of Dyslexia*, **57** (1), 3–32.

Vaessen, A., Gerretsen, P., & Blomert, L. (2009). Naming problems do not reflect a second independent core deficit in dyslexia: double deficits explored. *Journal of Experimental Child Psychology*, **103** (2), 202–221.

van Bergen, E., van der Leij, A., & de Jong, P.F. (2014). The intergenerational multiple deficit model and the case of dyslexia. *Frontiers in Human Neuroscience*, **8**, 346.

Vaughn, S., Wexler, J., Leroux, A., Roberts, G., Denton, C., et al. (2012). Effects of intensive reading intervention for eighth-grade students with persistently inadequate response to intervention. *Journal of Learning Disabilities*, **45** (6), 515–525.

Vellutino, F.R., Scanlon, D.M., Sipay, E.R., Small, S.G., Pratt, A., 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.

Vermunt, J.K., & Magidson, J. (2013). *Technical guide for Latent GOLD 5.0: Basic, advanced, and syntax*. Belmont, MA: Statistical Innovations Inc.

Vukovic, R.K., & Siegel, L.S. (2006). The double-deficit hypothesis: a comprehensive analysis of the evidence. *Journal of Learning Disabilities*, **39** (1), 25–47.

Waesche, J.S.B., Schatschneider, C., Maner, J., Ahmed, Y., & Wagner, R. (2011). Examining agreement and longitudinal stability among traditional and RTI-based definitions of reading disability using the affected-status agreement statistic. *Journal of Learning Disabilities*, **44** (3), 296–307.

Wagner, R.K., & Barker, T.A. (1994). The development of orthographic processing ability. In V.W. Berninger (Ed.), *The varieties of orthographic knowledge I: Theoretical and developmental issues* (pp. 243–276). Dordrecht: Kluwer Academic Publishers.

Wagner, R.K., Torgesen, J.K., & Rashotte, C.A. (1994). Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities*, **27** (5), 276–286.

Wagner, R.K., Torgesen, J.K., & Rashotte, C.A. (1999). *CTOPP: Comprehensive Test of Phonological Processing*. Austin, TX: PRO-ED.

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review*, **36** (4), 541–561.

Wanzek, J., Vaughn, S., Scammacca, N.K., Metz, K., Murray, C.S., et al. (2013). Extensive reading interventions for students with reading difficulties after grade 3. *Review of Educational Research*, **83** (2), 163–195.

Warmington, M., & Hulme, C. (2012). Phoneme awareness, visual-verbal paired-associate learning, and rapid automatized naming as predictors of individual differences in reading ability. *Scientific Studies of Reading*, **16** (1), 45–62.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children-WISC-IV.* San Antonio, TX: Psychological Corporation.

Wiederholt, J., & Bryant, B. (2012). *Gray oral reading test-(GORT-4)*. Austin, TX: PRO-ED.

Wimmer, H., Mayringer, H., & Landerl, K. (2000). The double-deficit hypothesis and difficulties in learning to read: a regular orthography. *Journal of Educational Psychology*, **92** (4), 668–680.

Wolf, M., Bally, H., & Morris, R. (1986). Automaticity, retrieval processes, and reading: a longitudinal study in average and impaired readers. *Child Development*, **57** (4), 988–1000.

Wolf, M., & Bowers, P.G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, **91** (3), 415–438.

Wolf, M., & Bowers, P.G. (2000). Naming-speed processes and developmental reading disabilities. *Journal of Learning Disabilities*, **33**, 322–324.

Wolf, M., & Denckla, M.B. (2005). *RAN/RAS: Rapid Automatized Naming and Rapid Alternating*. Austin, TX: PRO-ED.

Wolff, U. (2010). Subgrouping of readers based on performance measures: a latent profile analysis. *Reading and Writing*, **23** (2), 209–238.

Woodcock, R. (1998). *Woodcock Reading Mastery Tests-Revised (WRMT-R/NU)*. Circle Pines, MN: American Guidance Service.

Woodcock, R. (2011). *Woodcock Reading Mastery Test (WRMT-III)*. San Antonio, TX: Pearson.

Received: 15 February 2016
Accepted: 2 June 2016