Pre-literacy Heterogeneity in Dutch-speaking Kindergartners: Latent Profile Analysis

Cara Verwimp[1], Femke Vanden Bempt[1,2], Silke Kellens[1], Maria Economou[1,2], Maaike Vandermosten[2], Jan Wouters[2], Pol Ghesquière[1] & Jolijn Vanderauwera[1,2,3,4]

Affiliations:

[1] Parenting and Special Education Research Unit, Faculty of Psychology and Educational Sciences, KU Leuven, Belgium

[2] Research Group ExpORL, Department of Neurosciences, KU Leuven, Belgium

[3] Psychological Sciences Research Institute, Université Catholique de Louvain, Belgium

[4] Institute of Neuroscience, Université Catholique de Louvain, Belgium

ORCID iDs

Cara Verwimp https://orcid.org/0000-0002-5444-303X

Maria Economou https://orcid.org/0000-0001-8199-0234

Maaike Vandermosten https://orcid.org/0000-0002-9928-1580

Jan Wouters https://orcid.org/0000-0002-0093-698X

Pol Ghesquière https://orcid.org/0000-0001-9056-7550

Jolijn Vanderauwera https://orcid.org/0000-0002-7106-8534

Corresponding author: Jolijn Vanderauwera, Place Mercier 10, 1348 Louvain-la-Neuve, email: jolijn.vanderauwera@uclouvain.be

Declarations

Funding. This study was funded by the Research Council of KU Leuven (Grant C14/17/046). Jolijn Vanderauwera was a postdoctoral fellow of the Research Foundation - Flanders (FWO) (12T4818N).

Conflicts of interest. The authors report no conflicts of interest.

Availability of data and material. Data can be made available upon request and after compliance with EC regulations.

Authors' contributions. FVB, SK, ME, MV, JW, PG and JV were involved in the conception and design of the study. CV, FVB, and SK collected the data. CV performed the analyses and wrote the first version of the manuscript. All authors commented on the manuscript and approved the final version.

Acknowledgments. Special thanks to all children and their parents who kindly participated in this study and the students who helped with data collection.

Abstract

Research demonstrated that a dyslexia diagnosis is mainly given after the most effective time for intervention has passed, referred to as the dyslexia paradox. Although some pre-reading cognitive measures have been found to be strong predictors of early literacy acquisition, i.e. phonological awareness (PA), letter knowledge (LK) and rapid automatized naming (RAN), more insight in the variability of pre-reading profiles might be of great importance for early identification of children who have an elevated risk for developing dyslexia and to provide tailor-made interventions. To address this issue, this study used a Latent Profile Analysis (LPA) to disentangle different pre-reading profiles in a sample of 1091 Dutch-speaking kindergartners. Four profiles emerged: high performers (16.50%), average performers (40.24%), below-average performers with average IQ (25.57%) and below-average performers with below-average IQ (17.69%). These results suggested two at-risk profiles diverging in IQ, which are presumably more likely to develop dyslexia later on. Although below-average profiles differed significantly in rapid naming and IQ, no clear evidence for the double-deficit theory was found in Dutch-speaking kindergartners. Educational level and reading history of the parents appeared to be predictive for children's classification membership. Our results point towards the heterogeneity that is already present in kindergartners and the possibility to identify at-risk profiles prior to reading instruction, which may be the foundation for earlier targeted interventions. However, more extended research is needed to determine the stability of these profiles across time and across different languages.

Keywords: dyslexia, latent profile analysis, pre-reading profiles, heterogeneity

Learning to read is an important objective of primary education. Three to 7 % of the children is diagnosed with developmental dyslexia (Gersons-Wolfensberger & Ruijssenaars, 1997), a learning disability characterized by difficulties with learning to read and/or spell words, which could not be explained by intelligence and sensory abilities (Peterson & Pennington, 2012). However, dyslexia is typically diagnosed when the most effective time for intervention has passed and reading impairments show up to be severe and persistent (Wanzek & Vaughn, 2007), referred to as the 'dyslexia paradox' (Ozernov-Palchik & Gaab, 2016). Due to this late diagnosis, children often experience negative consequences such as depressive thoughts, feelings of shame and loneliness (Mugnaini, Lassi, La Malfa, & Albertini, 2009; Willcutt & Pennington, 2000). Additionally, children with reading impairments receive fewer opportunities to develop reading comprehension since they may be less exposed to written text, resulting in a large gap with peers (Ferrer et al., 2015). Therefore, early identification of reading problems is important to prevent and ameliorate socio-emotional problems and to provide effective interventions during early stages of literacy acquisition.

Several cognitive pre-reading measures have been identified as strong predictors of reading development across different languages, including phonological awareness (PA), letter knowledge (LK), and rapid automatized naming (RAN) (Boets et al., 2010; Caravolas et al., 2012; Georgiou, Torppa, Manolitsis, Lyytinen, & Parrila, 2012). Phonological awareness reflects the ability to perceive and manipulate individual sounds in spoken words (Gellert & Elbro, 2017). A meta-analysis of Melby-Lervåg, Lyster and Hulme (2012) showed that PA in kindergarten significantly correlated with reading development during the first years of reading instruction and is often impaired in children with dyslexia compared to age-matched controls. Letter knowledge, the knowledge of letter names, their sound and visual representation, is rapidly mastered beyond kindergarten and therefore has been found to be the most predictive prior to formal reading instruction (Schatschneider et al., 2004). Rapid automatized naming, the ability to quickly name series of familiar items (e.g. colors, objects, numbers or letters), has also been found as predictive for later reading performance (Araujo, Reis, Petersson, & Faisca, 2015).

Although these predictors are reliably found in several studies, predicting which children truly develop dyslexia prior to formal reading instruction remains challenging, as accurate prediction involves a trade-off between sensitivity (i.e. identifying children who are at risk to develop reading difficulties as positive) and specificity (i.e. identifying children without a cognitive risk to develop reading difficulties as negative). None of these predictors seem to be suitable for high sensitivity and specificity on its own (Ozernov-Palchik & Gaab, 2016; Pennington et al., 2012), often resulting in high rates of over- or under-identification. False identification in children without a risk (false positives) may prompt unnecessary concern for caregivers and unnecessary use of intervention resources (Poulsen, Nielsen, Juul, & Elbro, 2017). However, overlooking children with an elevated risk to develop dyslexia (false negatives), particularly those who go on to develop dyslexia, may lead to long-term consequences, as they cannot benefit from interventions at early age (Mugnaini et al., 2009; Willcutt & Pennington, 2000). As the percentage of false positives and negatives varies with the accuracy of the screening instrument and the selected cut-off score, selecting the most optimal cut-off is challenging, as one wants to avoid missing children who develop dyslexia at later age.

A first step towards earlier intervention is clarifying the behavioral heterogeneity within the reading population as different underlying patterns of impairment have been reported in dyslexia (Ozernov-Palchik & Gaab, 2016). For a long time, researchers tried to reduce the variety of behavioral symptoms to one single underlying cause, such as a phonological deficit (Pennington, 2006). However, it is now recognized that a phonological deficit does not always result in reading impairment (Boets, Wouters, van Wieringen, De Smedt, & Ghesquière, 2008).

Moreover, not all individuals who are reading impaired suffer from a phonological deficit (Wolf, O'Rourke, Lovett, Cirino, & Morris, 2002). As evidence showed that many impaired readers represent naming-speed deficits, Wolf and Bowers (1999) stated that both components, PA and RAN, represent distinct sources of reading dysfunctions which can be affected independently. In this case, three subtypes of reading impairment may appear: a PA deficit subtype, a RAN deficit subtype or a combined deficit of PA and RAN. This latter subtype may characterize individuals with the most profound reading impairment compared to the single deficit subtypes (Wolf & Bowers, 1999). However, this double-deficit hypothesis (DDH) claims that having either a rapid naming or phonological deficit is sufficient to develop dyslexia, although research showed that children with a persisting single deficit could be spared from a reading disability (Snowling, Bishop, & Stothard, 2000). Therefore, Pennington (2006) proposed a multiple deficit model in which dyslexia is caused by multiple cognitive factors which are likely to interact with risk and protective factors at different levels (i.e. genetic, environmental and neurobiological). These factors and their interactions might alter behavioral outcomes in dyslexia, resulting in a continuously distributed liability for reading failure, in which no single etiological factor is sufficient to cause dyslexia (Pennington, 2006; van Bergen, van der Leij, & de Jong, 2014). Moreover, the intergenerational multiple deficit model emphasizes the importance of parental influences on the offspring's liability to develop dyslexia (van Bergen et al., 2014). Hence, the behavioral outcome can be influenced by a large variability of factors, which may result in struggling readers with each their own strengths and weaknesses. Different behavioral outcomes might need different individualized intervention programs. Therefore, addressing this heterogeneity within the (pre-)reading population is of great importance to provide well-adapted interventions.

Although defining children at risk for dyslexia in kindergarten is currently not a prevailing approach in clinical practice, for research purposes these children are generally identified based on employing a cut-off score on one or more continuous cognitive measures, such as phonological awareness. However, studies tend to use different cut-off criteria on a different cognitive predictor to characterize risk, leading to inconsistency of findings. Moreover, according to the multiple-deficit model, a combination of predictors should be used rather than the performance on a single measure. This approach thus yields artificial subgroups as for example, children scoring close to the threshold can be classified differently despite their similar performance. Besides, substantial heterogeneity within at-risk and non-risk groups is likely. To address this issue, it is worth exploring other methods to identify children within the pre-reading population who present an elevated risk for developing dyslexia later on.

Techniques such as cluster analysis or latent profile analysis (LPA) deliver detailed and quantitative insights from large samples (e.g. Kornilov & Grigorenko, 2018; Oberski, 2016; Ozernov-Palchik et al., 2017), without being influenced by previous assumptions regarding the number of reading subtypes and pre-defined cut-off scores. Once clusters are identified, individuals can be assigned to the profile for which the posterior probability is the highest, providing a way to disentangle profiles with different cognitive deficits. The LPA technique allows to implement an individual's overall pattern of scores, making it an excellent tool to study complex multidimensional phenomena such as dyslexia (Lanza & Cooper, 2016). However, despite these benefits, concerns are raised by critics. First, many applications of LPA are exploratory. Which variables or criteria to use to obtain the final model are decided on a study level, making it challenging to compare results of different studies. Additionally, as dyslexia is multifactorial (Pennington, 2006), one may consider including a variety of variables as well as to implement co-occurring disorders such as ADHD, depression or anxiety disorder (Osman, 2000). However, including more parameters generally results in a better model fit, but may be less valuable for out-of-sample prediction. As a result, LPA is especially useful in identifying subgroups that are widely represented within the general sample whereas combinations that constitute a small proportion are difficult to detect. Arguably, one might be able to find smaller subgroups when using cut-off scores on separate variables. However, a combination of factors operates probabilistically by increasing (or decreasing) the risk that one develops a reading disorder (Pennington, 2006; van Bergen et al., 2014), pointing towards the need of a new approach in which different variables can be included concurrently. Therefore, we argue that LPA can offer an important contribution to the enhancement of early, personalized interventions, although it should be applied in combination with clinical judgement.

LPA has been used in previous research to disentangle profiles based on reading performance measures. Wolff (2010) identified eight profiles in a Swedish reading sample of 9-year-old children, based on word reading, reading speed, reading comprehension, and interpretation of documents in which children had to look for and process selected facts rather than read every word: (1) high performers, (2) average performers, (3) poor document 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, (7) low-average decoding, low-average fluency, low comprehension and (8) low decoding, poor fluency, average comprehension (dyslexic). Another study found five profiles based on word recognition and reading comprehension in a Finnish sample (first and second graders): (1) poor readers, (2) slow decoders, (3) poor comprehenders, (4) average readers and (5) good readers (Torppa et al., 2007). The authors reported differences among profiles in early language and literacy skill development as well as in reading experiences. Ozernov-Palchik et al. (2017) applied an LPA in an English-speaking sample of 1215 (pre-)kindergartners, of which 30.6% was considered reader. Six profiles emerged based on performance on PA, LK, RAN and verbal short-term memory (vSTM): (1) average performers, (2) below-average performers, (3) high performers, (4) PA risk, (5) RAN risk and (6) double-deficit risk (both RAN and PA). Using a latent class regression model (Vermunt & Magidson, 2005), the authors observed that profile membership in kindergarten could accurately predict the reading performance at the end of the first grade, as specific patterns of performance remained stable across time (Ozernov-Palchik et al., 2017).

Although several studies already used LPA, the previously mentioned studies used samples of children that already started with reading instruction to some extent. However, referring to the 'dyslexia paradox', individualized interventions should take place prior to formal reading instruction to improve effects (Catts, Nielsen, Bridges, Liu, & Bontempo, 2015; Wanzek & Vaughn, 2007). Therefore, it is of great importance to identify which profiles are present within the pre-reading population based on cognitive variables predictive for later reading acquisition. More specifically, the aim of the present study is not only to investigate whether a cognitive risk group to develop dyslexia could be identified prior to reading instruction, but also to investigate whether within this at-risk group, different profiles appear that point towards different deficits. Additionally, as English has no consistent grapheme-phoneme correspondences (Schmalz, Marinus, Coltheart, & Castles, 2015; Share, 2008), this study aimed to extend the aforementioned study of Ozernov-Palchik et al., (2017) to a Dutch-speaking sample of kindergartners to investigate whether similar (at-risk) profiles appear in a more transparent language (i.e. Dutch). Three cognitive measures of early literacy, i.e. PA, LK and RAN, and non-verbal intelligence, were assessed in third grade of kindergarten to estimate the number of profiles. In sum, two novel features were that in the Flemish school system, all children were non-reading at the moment of testing and the highly consistent spelling-to-sound language, i.e. relatively consistent in mapping graphemes into phonemes, in Dutch language compared to English (Schmalz et al., 2015; Seymour, Aro, & Erskine, 2003), which may result in different profiles. Second, according to the multiple-deficit model of Pennington (2006), we investigated whether socio-economic status and reading history of both parents appeared to be predictive for the child's membership classification. To ensure parsimony of the model and to disentangle child and parental factors, only child factors were included in the LPA and parental influences were examined afterwards.

Method

Participants and Procedure

Participants were recruited prior to the third year of kindergarten by sending an information video and information letters to parents in schools throughout Flanders (Belgium). Recruitment took place within the framework of a broader research project. Out of the 1900 parents and children who gave consent to be contacted, 1225 children were selected based on questionnaires, ensuring that all included children were native Dutch speakers with no history of brain damage, vision deficits, articulatory or language problems or any gross hearing deficiencies. Cognitive measures were administered prior to formal reading instruction, at the start of the last year of kindergarten, within a two-month time span. At this time, in line with Flemish legislation (http://www.ond.vlaanderen.be/), none of the children received formal reading instruction yet and should be considered as pre-reader.

Children completed an assessment battery comprising three blocks: two tablet sessions and one paper-pencil session, with breaks in between. For both tablet sessions, a maximum of eight children were positioned in a quiet room at the children's school with a Samsung Galaxy Tab E 9.6 tablet and a headphone (Audiotechnica ATH-M20x) for each child. Cardboard screens were placed between the children to facilitate concentration. The tablet sessions consisted of a digital version of Raven's Coloured Progressive Matrices (CPM; Raven, Court, & Raven, 1984) to assess non-verbal intelligence and an adapted version of the tablet game Diesel-X, comprising three mini-games: beginning phoneme identification, end phoneme identification and receptive LK. This game was developed for preschoolers to assess cognitive risk to develop dyslexia and originally included a frequency modulation task instead of a beginning phoneme identification task (Geurts et al., 2015). The paper-pencil session was conducted individually and comprised three tasks: RAN, productive LK and end rhyme identification.

Children with incomplete or unreliable scores (n = 21) and children with low non-verbal IQ (n = 113) were excluded from further analyses. Given that no normative data was available for the digital version of Raven CPM, low non-verbal IQ was defined by the lowest 10 percent of scores (*M* = 98.15, *SD* = 16.66). The final sample included 1091 children (520 boys), with an average age of 5 years and 3 months (*SD* = 0.28). This study was approved by the ethics committee of KU Leuven and written parental consent for the participation of the children was obtained in line with the Declaration of Helsinki.

Assessment Battery

Cognitive measures. Three cognitive components were assessed and quantified by means of factor scores to control for different difficulties in tasks within one component. The first component was PA, based on three tasks: beginning phoneme, end phoneme and end rhyme identification. In the first task, all items were visually and auditory presented via the tablet game Diesel-X (Geurts et al., 2015). Children had to indicate the item with the same beginning phoneme as the target item. Each item consisted of one reference item and four alternatives. Distracter alternatives were systematically constructed to prevent guessing. One item was semantically related (e.g. 'cat' and 'dog' or Dutch alternative: 'boom' and 'tak'), one was phonologically related (e.g. 'cat' and 'map' or Dutch alternative: 'boom' and 'doos') and one item was not related to the reference item (e.g. 'cat' and 'bed' or Dutch alternative: 'boom' and 'wieg'). Diesel-X provided two practice items prior to the test items where feedback was given to ensure that the children were familiar with the task. Second, children completed an end phoneme identification task, similar to the first task but in which children had to indicate the item with the same end phoneme instead of the same beginning phoneme (Geurts et al., 2015). Both tasks consisted of two practice items and 10 test items and the score was calculated as the number of correct responses independent of time, with a maximum score of 10. Finally, a paper-pencil rhyme task was administered. Children had to indicate the item with the same end rhyme as the target item (Boets et al., 2010). Each item was visually presented, followed by a row of four pictures including a correct, semantically related, phonologically related and non-related answer. All items, high frequent one-syllable Dutch words, were named for the children. The task consisted of 12 items, preceded by two practice items. The raw score was computed as the number correctly answered items, with a maximum score of 12.

A LK factor score was computed based on a productive and receptive subtest, assessed by a paper-pencil task and a Diesel-X task, respectively. For the productive subtest, children had to name the 16 most common letters in Dutch language that were visually presented (Boets, et al., 2008). Both letter names and sounds were considered correct. In the receptive letter knowledge task, a letter was named and children had to select the correct letter out of five alternatives (Geurts et al., 2015). The raw score in both tasks was the number of letters that was named correctly, with a maximum score of 16.

A RAN factor score was computed based on a subtest for colors and a subtest for objects (van den Bos, Zijlstra, & Spelberg, 2002). Children had to name an array of 50 familiar colors or objects, presented in 5 columns of 10 stimuli, as quick and accurately as possible (Boets et al., 2010). The colors were represented by small colorized rectangles in five colors: black, blue, red, yellow and green. The objects represented five high-frequent words: tree ('boom'), chair ('stoel'), scissors ('schaar'), bicycle ('fiets') and duck ('eend'). Prior to the start of the task, children had a practice session in which they named the last column to ensure that they were familiar with instructions and names of colors and objects. The raw score was defined as the number of accurately named items divided by the time that was needed to complete the test.

Parental and environmental measures. To assess parental and environmental factors, i.e. adult reading history and educational level as a measure of socio-economic status (SES), parents had to fill in a Dutch translation of the Adult Reading History Questionnaire (ARHQ; Lefly & Pennington, 2000) and a general self-constructed questionnaire. The ARHQ is a self-report screening tool, which exists of 26 items to measure family history of reading problems (Lefly & Pennington, 2000). Each item consists of a Likert scale with scores ranging from 0 to 4, with lower scores indicating less difficulty in reading history and current reading habits. Scores were computed by dividing the total score on the 23 first items by the maximum score of the questionnaire (92). The ARHQ has been filled out by 941 mothers and 693 fathers and was included as a continuous variable in analyses. SES was defined by the educational level of the parents using three categories. Parents without a higher education degree received a score of 1, parents with a higher education bachelor's degree received a score of 2 and parents with a master's or PhD degree received a score of 3. This questionnaire has been filled out by 1040 mothers and 1006 fathers and was included as a categorical variable in analyses.

Data Analysis

To identify homogeneous profiles based on reading-related cognitive variables in kindergarten (i.e. PA, LK and RAN), a latent profile analysis was conducted in Latent Gold 5.1 (Vermunt & Magidson, 2005). The model-fitting started with a one profile model which hypothesizes that all children belong to one homogenous group. Subsequently, additional profiles were added one at the time in order to obtain the best fitting model given the data. In this study, 1-6 profile models were fit to the data. A combination of criteria was used to determine the optimal number of classes, which are further explained below: (1) information criteria (Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)); (2) conditional bootstrap (Bootstrap -2LL Diff) based on Log-Likelihood (LL); (3) bivariate residuals; (4) cluster interpretation; (5) classification error; and (6) magnitude of R^2 values (Vermunt & Magidson, 2005). First, information criteria (i.e. BIC and AIC) were used to avoid overfitting. Specifically, these criteria adjust the LL in order to account for the number of estimated parameters in the model, with lower values indicating a better fit of the model to the data (Vermunt & Magidson, 2005). Second, due to the continuous dataset, the common approach of selecting the model based on the *p*-value associated with the L^2 fit statistic from a Chi-squared table could not be used (Vermunt & Magidson, 2005). Therefore, in order to test whether additional classes improved the amount of fit, the difference in log-likelihood value between models with k (H_1) and k-1 (H_0) classes was used to estimate a bootstrapped *p*-value (Vermunt & Magidson, 2016). A significant *p*-value indicated a significant improvement in model fit and therefore the more restricted model was rejected in favor of the less restricted model. Third, bivariate residuals substantially larger than 2.0 suggested local dependencies which could not be adequately explained by the model (Vermunt & Magidson, 2005). If this was the case, relationships were set as direct effects and models including direct effect parameters were compared to their restricted model in order to test whether including direct effects improved the model fit (Vermunt & Magidson, 2005). When the model did not improve in fit, the less restricted model was chosen as best solution for reasons of parsimony. Fourth, clusters should at least include 10% of participants and should have distinct cluster characteristics for interpretability (Dunn, Campbell, & Jordan, 2013; Vermunt & Magidson, 2005). Therefore, models with clusters comprising less than 10% of participants were dropped from consideration. Fifth, different models were compared according to the classification error, i.e. the proportion of cases that are estimated to be misclassified. Finally, R^2 values indicated how well an indicator was explained by the model, similar to item communalities in factor analysis (Vermunt & Magidson, 2005).

Once the latent profiles were identified, follow-up analyses were conducted in the statistical software package R (version 1.1.4). To check for differences among the profiles in pre-reading measures and ARHQ, a univariate analysis of variance (ANOVA) was performed. Chi-square tests were conducted to evaluate the relationship between SES and class membership. While the LPA was conducted on the whole sample (N = 1091), some children were excluded from follow-up analyses when parental questionnaires were missing. Significant effects in follow-up analyses were further investigated with pairwise post hoc comparisons, applying False Discovery Rate (FDR) correction (Benjamini & Hochberg, 1995). For both LPA and follow-up analyses, the significance level was set at $p \leq .01$ in order to account for the higher number of false positives in large samples.

Results

Latent Profile Analysis

The goodness of fit statistics are shown in Table 1. The baseline model that estimated a 1-profile model was used to test whether subsequent models with additional profiles provided a significant improvement in model fit. As shown in Table 1, the BIC value decreased from one to four classes, whereas the AIC value decreased from one up to six classes. This indicated three candidates for best fit, i.e., the 4-, 5-, and 6-cluster model. Therefore, other information was used to determine the best fitting model. The log-likelihood -2LL was computed for these models to test which model increased in fit compared to its restricted model. The 4-cluster model provided more information than the 3-cluster model ($-2LL_{Diff} = 78.968$, $p < .001$) and the 5-cluster model was superior to the 4-cluster model ($-2LL_{Diff} = 53.058$, $p < .001$) but for the latter the bivariate residuals still exceeded the recommended value of 2.0 (Vermunt & Magidson, 2005). In order to minimize these residual values, a 6-cluster model was estimated. By adding another latent class, bivariate residuals decreased but the log-likelihood -2LL showed that the 6-cluster model was not superior to the 5-cluster model ($-2LL_{Diff} = 21.278$, $p = .106$). Considering the 5-cluster model, one cluster only included 6% of participants and had an increased estimated classification error compared to the 4-cluster model (21.8% versus 18.8%). Therefore, the 5-cluster model was discarded in favor of the 4-cluster model.

In order to deal with bivariate residuals that substantially exceeded 2.0 in the 4-cluster model, direct effects were added to account for residual correlation (Vermunt & Magidson, 2016). Models with direct effects were tested against their restricted models using the conditional bootstrap (Bootstrap -2LL Diff). This showed that inclusion of one direct effect (IQ and PA), provided no significant improvement over the traditional 4-cluster model ($-2LL_{Diff} = 14.173$, $p = .016$). A model with two direct effects (IQ and PA; LK and PA) was significantly superior to the restricted 4-cluster model ($-2LL_{Diff} = 24.798$, $p = .004$) but not significantly superior to the 4-cluster model with one direct effect ($-2LL_{Diff} = 10.625$, $p = .014$). In addition, the classification error increased by adding direct effects. Therefore, the traditional 4-cluster model was chosen as the main model for reasons of parsimony since adding effects did not significantly improve the fit. This model is further discussed in the next section.

Table 1 *Goodness of fit statistics for model comparisons.*

Model	LL	BIC(LL)	AIC(LL)	AIC3(LL)	CAIC(LL)	Npar	Class.Err	$-2LL_{Diff}$
1-cluster	-9121.494	18302.948	18262.989	18270.989	18310.948	8	0.000	
2-cluster	-8509.112	17137.138	17052.225	17069.225	17154.138	17	0.059	1228.764*
3-cluster	-8413.503	17008.872	16879.006	16905.006	17034.872	26	0.144	191.219*
4-cluster	-8374.019	**16992.858**	16818.038	16853.038	**17027.858**	35	0.188	78.968*
5-cluster	-8347.490	17002.753	16782.980	**16826.980**	17046.753	44	0.218	53.058*
6-cluster	-8336.851	17044.429	**16779.701**	16832.701	17097.429	53	0.221	21.278

Note: Lowest values are represented in bold. LL = log likelihood; BIC = Bayesian Information Criterion; AIC = Akaike Information Criterion; AIC3 = Corrected AIC with penalty factor of three; CAIC = Consistent AIC; Npar = Number of estimated parameters; Class. Err. = Classification error. Adjacent models were compared using the Bootstrapped Likelihood Ratio Test ($-2LL_{Diff}$). *p < .01

Profile descriptions. Table 2 shows the estimated parameters for each of the four profiles (see Figure 1). This table demonstrates that each factor was associated with different levels across classes, showed by the significant Wald tests. Additionally, each predictor was sufficiently explained by the model, showed by the R^2 values ranging from 20.36% for RAN to 72.26% for LK. The beta parameter indicates the strength of the effect of the predictor on the class (with $-2.33 < z < 2.33$ indicating significance on an alpha level of .01). The profiles are ordered by performance of group members, from highest performance to lowest. Individual cases were assigned to the profile for which the posterior probability was the highest. Estimated sizes of the profiles based on posterior membership probabilities were not exactly reproduced by using the modal assignment rule, resulting in a classification error of 18.77%.

Profile 1. The smallest profile included 180 children (16.50%). Table 2 shows that this profile is positively influenced by IQ ($\beta = 10.29$; $z = 9.38$), PA ($\beta = 1.36$; $z = 18.69$), LK ($\beta = 1.66$; $z = 20.08$) and RAN ($\beta = 0.62$; $z = 9.15$). Therefore, higher scores on these three pre-reading measures and IQ were associated with a greater probability of belonging to this profile, which we may call the '*high performers* (*HP*)'.

Profile 2. The largest group, referred to as '*average performers (AP)*', included 439 children (40.24%). This profile is associated with an average non-verbal IQ ($\beta = 3.06$; $z = 3.42$), performance near the mean on phonological awareness ($\beta = -0.01$; $z = -0.19$) and rapid naming ($\beta = 0.30$; $z = 4.36$) and a slightly lower score on letter knowledge ($\beta = -0.12$; $z = -3.11$). In other words, average scores on non-verbal IQ, PA and RAN resulted in a greater probability of belonging to this profile.

Profile 3. The *'below-average performers with average IQ (BP with AIQ)*' profile included 279 children (25.57%), with an average non-verbal IQ ($\beta = 2.92$; $z = 2.67$). This profile is negatively influenced by PA ($\beta = -0.64$; $z = -11.55$), RAN ($\beta = -0.55$; $z = -7.44$) and LK ($\beta = -0.73$; $z = -16.77$). Performing below the mean across all pre-reading measures is therefore associated with a greater probability of belonging to this profile.

Profile 4. The *'below-average performers with below-average IQ (BP with BIQ)*' profile included 193 children (17.69%). This profile is negatively influenced by PA ($\beta = -0.71$; $z = -9.99$), LK ($\beta = -0.81$; $z = -11.96$) and RAN ($\beta = -0.37$; $z = -4.40$) and thus had a similar pattern of performance as the previous profile. The greatest difference compared to the previous profile is the negative influence of non-verbal IQ ($\beta = -16.27$; $z = -25.02$). This means that children with low scores on the three pre-reading measures and a lower IQ score were more likely to belong to this profile, whereas the children with an average IQ score were more likely to belong to the previous profile.

Table 2 *Estimated parameters for the four-cluster model.*

Cluster	High performers		Average performers		Below-average performers with average IQ		Below-average performers with below-average IQ			
Parameter	β	z	β	z	β	z	β	z	Wald	R^2
Raven IQ	10.29	9.38	3.06	3.42	2.92	2.67	-16.27	-25.02	637.14*	0.29
Factor PA	1.36	18.68	-0.01	-0.19	-0.64	-11.55	-0.71	-9.99	360.74*	0.50
Factor LK	1.66	20.08	-0.12	-3.11	-0.73	-16.77	-0.81	-11.96	446.06*	0.72
Factor RAN	0.62	9.15	0.30	4.36	-0.55	-7.44	-0.37	-4.40	165.13*	0.20

Note: *$p < .001$.

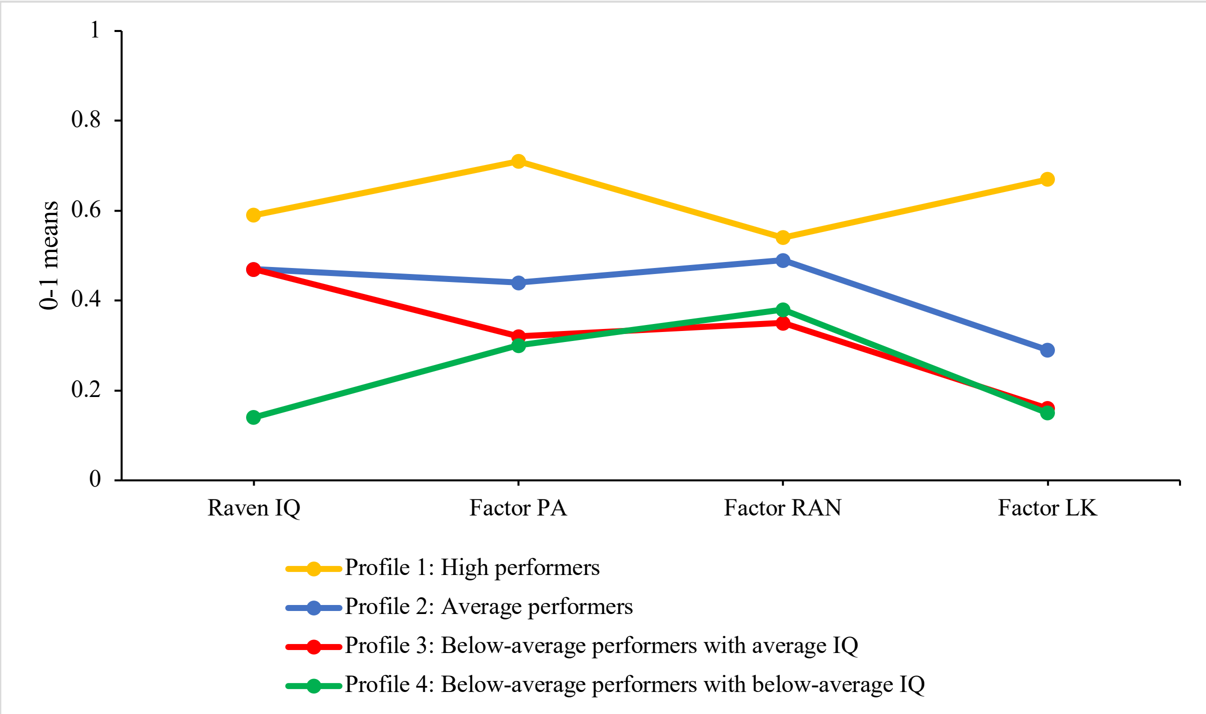

Figure 1. Profile plot for the optimal four-cluster solution. By default, class-specific means were rescaled by Latent Gold by subtracting the lowest observed value from the class-specific means and dividing by the range to lie within a 0-1 range.

A one-way ANOVA confirmed significant differences in pre-reading measures among the profiles ($p_s < .010$; Table 3). Post hoc tests revealed that RAN scores differed significantly in all profiles ($p_s < .007$), IQ scores differed in all profiles ($p_s < .010$) except between *AP* (Profile 2) and *BP with AIQ* (Profile 3) ($p = .030$), and LK and PA scores differed in all profiles

(p_s < .001) except between *BP with AIQ* and *BIQ* (Profile 3 and 4) (p = .390 and p = .360 for LK and PA, respectively). All comparisons survived the FDR correction.

Table 3 *Means (and standard deviations) and ranges for all measures of the final sample (N = 1091).*

		Range	**HP**	**AP**	**BP with AIQ**	**BP with BIQ**	**ANOVA (*F*)**
Age (Y;M)		4;8–5;10	5.30 (0.29)	5.25 (0.28)	5.18 (0.26)	5.17 (0.28)	
Non-verbal IQ		75.38–133.75	109.72 (15.30)	102.94 (13.35)	104.93 (10.66)	82.87 (4.11)	$F_{(3,1087)} = 194.451$*
PA	Factor score	-1.41–3.29	1.48 (0.77)	0.09 (0.75)	-0.62 (0.57)	-0.68 (0.54)	$F_{(3,1087)} = 432.160$*
	Beginning phoneme	0–10	8.12 (2.07)	4.96 (2.21)	3.21 (1.65)	3.19 (1.50)	
	End phoneme	0–10	5.88 (2.50)	2.88 (1.73)	2.02 (1.28)	1.99 (1.20)	
	End rhyme	0–12	10.51 (1.73)	8.97 (2.45)	7.23 (2.95)	6.86 (2.99)	
LK	Factor score	-1.41–3.29	1.83 (0.77)	0.01 (0.48)	-0.69 (0.28)	-0.73 (0.26)	$F_{(3,1087)} = 1236.150$*
	Productive LK	0–16	11.08 (3.33)	3.99 (2.33)	1.16 (1.18)	0.99 (1.15)	
	Receptive LK	0–16	10.96 (2.82)	5.51 (1.91)	3.47 (1.41)	3.38 (1.28)	
RAN	Factor score	-2.79–3.50	0.59 (0.95)	0.35 (0.93)	-0.64 (0.65)	-0.42 (0.93)	$F_{(3,1087)} = 117.285$*
	Colors	0.20–1.19	0.71 (0.16)	0.68 (0.149)	0.53 (0.11)	0.56 (0.15)	
	Objects	0.23–1.13	49.67 (0.14)	49.63 (0.14)	49.46 (0.10)	0.59 (0.14)	

Note: Subtests are shown as raw scores. For each pre-reading measure (PA, LK, RAN), a factor score was computed based on the subtests. Mean scores (based on factor scores) were compared across profiles by a one-way ANOVA. *$p < .01$

Parental and Environmental Factors

To assess the relationship between ARHQ and class membership, a one-way ANOVA was conducted with ARHQ as dependent variable and class membership as between subject factor. Only children who had ARHQ scores were included in these analyses (mothers: n = 941, fathers: n = 693). This revealed a significant effect of ARHQ among profiles for both mothers ($F_{(3,937)} = 5.278$, $p = .001$) and fathers ($F_{(3,689)} = 8.836$, $p < .001$). Moreover, post hoc tests revealed that the ARHQ from the mothers in the *HP* (Profile 1) differed significantly from the two *BP* profiles (Profile 3 and 4) ($p_s < .001$). Similar results were found for the ARHQ from the fathers ($p_s < .001$). Furthermore, regarding the paternal ARHQ, both *BP* profiles (Profiles 3 and 4) differed from the *AP* (Profile 2) ($p_s < .007$). Effects surviving FDR correction for multiple comparisons are shown in Figure 2.

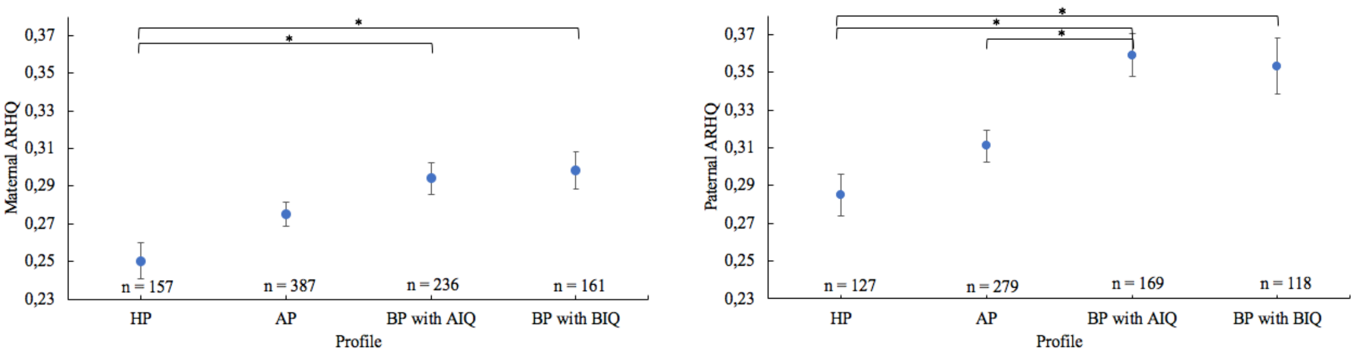

Figure 2. Differences between profile membership regarding maternal (left) and paternal (right) Adult Reading History Questionnaire. Lower values indicate less difficulties in reading. Standard errors are shown as error bars. *p < .01 (corrected p-values). HP = High performers, AP = Average performers, BP with AIQ = Below-average performers with average IQ, BP with BIQ = Below-average performers with below-average IQ.

To assess the relationship between SES and class membership (Figure 3), Chi-square tests were conducted. This revealed significant differences in profile distribution across the three levels of maternal education ($\chi^2_{(6)} = 32.921$, $p < .001$, $\phi_C = .126$). Figure 3 shows that the majority of *AP* (Profile 2) and *HP* (Profile 1) belonged to the medium (36.43% and 36.00% respectively) and high SES groups (47.19% and 52.57%, respectively). In contrast, only

16.38% of the *AP* and 11.43% of the *HP* had a mother with a low educational level. The majority of *BP with AIQ* (Profile 3) had a mother with middle or high SES (40.60% and 35.71% respectively) whereas *BP with BIQ* (Profile 4) were more equally divided over the three levels of maternal education (28.73% in low SES, 39.23% in middle SES and 32.04% in high SES). Looking at paternal education also revealed significant differences in profile distribution across the three SES groups ($\chi^2_{(6)} = 43.226$, $p < .001$, $\phi_C = .147$). *AP* were almost equally divided over the three levels of paternal education (32.91% in low SES, 30.87% in middle SES and 36.22% in high SES), whereas almost half of the *BP* had a father with a low education level (42.31% and 48.00% for the average and below-average IQ profiles respectively). Only 21.14% of the *BP with BIQ* had a father with a master's degree or PhD, contrary to the *HP*, of which half of the children had a father with a master's degree or PhD (50.29%).

To examine which relations mainly drove the association between SES and profile membership, Chi-square tests were conducted for each association separately (each educational level combined with each profile). Regarding maternal educational level, a significant effect was found between low SES and membership of *BP with BIQ* (Profile 4) and *HP* (Profile 1) ($p_s < .003$). Furthermore, a significant association was found between a high SES and membership of *BP with BIQ* (Profile 4) and *HP* (Profile 1) ($p_s < .003$). In other words, there were fewer children than expected who were classified as *HP* and had a mother with a low SES and fewer children than expected who were classified as *BP with BIQ* and had a mother with a high SES. In contrast, there were more children than expected categorized as *BP with BIQ* with a mother with a low SES and categorized as *HP* with a mother with a high SES. Similar results can be found based on the paternal SES. All comparisons survived the FDR correction for multiple comparisons.

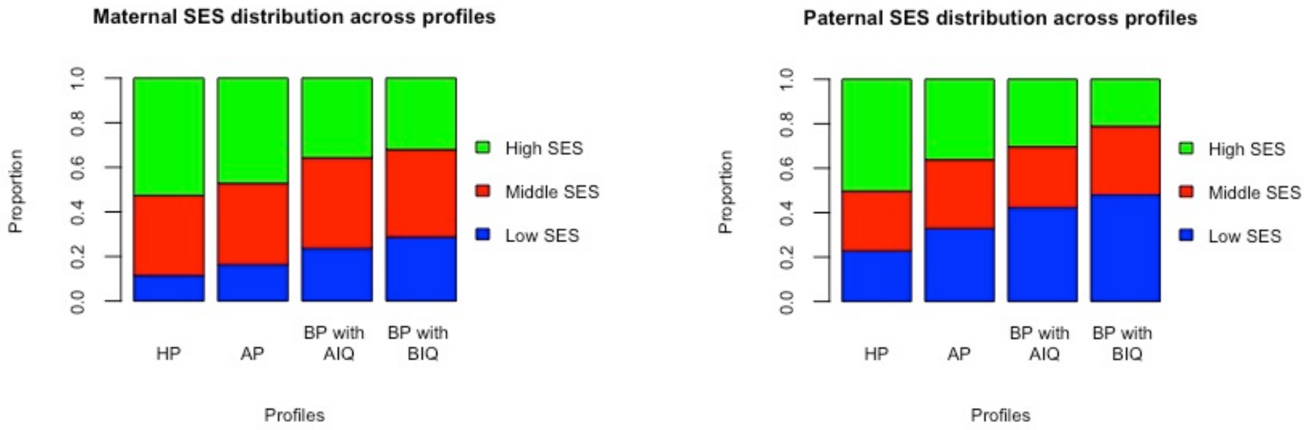

Figure 3. Maternal (left) and paternal (right) SES distribution across profiles. Note that proportion adds up to 1 within one profile. HP = High performers, AP = Average performers, BP with AIQ = Below-average performers with average IQ, BP with BIQ = Below-average performers with below-average IQ.

Discussion

A dyslexia diagnosis is typically given in second grade or later. However, earlier interventions are shown to be more effective, pointing to the importance of early identification (Ozernov-Palchik & Gaab, 2016; Wanzek & Vaughn, 2007). Although some pre-reading measures have been found as strong predictors of early literacy acquisition (i.e. PA, LK and RAN), a dichotomy between being or not being at risk does not provide information for targeted interventions. To address this issue, this study investigated whether heterogeneous profiles could be identified prior to formal reading instruction in 1091 Dutch-speaking kindergartners, using a Latent Profile Analysis. This way, long-term outcomes might be improved as interventions could be targeted towards individual deficits. The model included three typical pre-reading measures, PA, LK and RAN, and a non-verbal IQ score. Four distinct profiles emerged of which two could be identified as at-risk profiles.

These 4 profiles were: (1) high performers (16.50%); (2) average performers (40.24%); (3) below-average performers with average IQ (25.57%); and (4) below-average performers with below-average IQ (17.69%). All profiles were associated with different levels of scores, except the two latter, which had similar scores on PA and LK but differed significantly in RAN and non-verbal IQ scores. *The high performers* were associated with scores above the mean across all cognitive measures whereas the *average performers* had scores near the mean across all measures. The two below-average profiles were associated with scores below the mean, suggesting that two profiles emerged from our data, which are presumably both more likely to develop dyslexia. This corroborates Stanovich's (1996) claim that there are no significant differences in fundamental cognitive processes leading to reading difficulties in children with high and low intelligence scores. This reflects the fierce criticism against the discrepancy criterion, in which differences between IQ and achievement are used to identify learning disabilities. Similarly, the study of Ozernov-Palchik and colleagues (2017) did not provide evidence for the IQ-discrepancy model of dyslexia, as no pattern of low language skills despite an average IQ emerged. Specifically, they found that only the group of participants with a RAN-deficit was associated with an average IQ, whereas the PA-deficit group and double-deficit group had a low-average IQ. As IQ does not determine the development of dyslexia, we hypothesize that children in both below-average profiles have an elevated risk for developing dyslexia based on cognitive pre-reading skills and thus are both eligible for receiving earlier interventions. It is however possible that these profiles are in need of different interventions. For instance, *below-average performers with below-average IQ* might benefit more from a broader intervention, including non-verbal abilities whereas *below-average performers with average IQ* might only need intervention based on reading-related abilities.

Taking a more thorough look at the scores revealed that *average performers* with higher scores on the cognitive measures had a higher probability to belong to the *high performers* whereas children with lower scores had a higher probability to belong to the *below-average performers*. Moreover, children with two below-average scores and one high score on the third pre-reading measure, were classified as *average performer*. This is in line with the multiple deficit model of Pennington (2006), in which one factor can have a negative or positive influence on the outcome. This highlights the importance of taking the full profile of the child into account, rather than relying on a single deficit model, which indicates the applicability of using these profiles instead of an arbitrary cut-off score.

Direct comparisons between profiles found in previous research and the present study must be interpreted cautiously due to the variability in measures, sample characteristics (e.g. differences in age and reading experience) and differences in classification techniques. However, it is interesting to note that the profiles that were retrieved in the present study were partly in line with the profiles reported by Ozernov-Palchik et al (2017). On the one hand, we found similar profiles with an overall performance level, such as *high performers* and *average performers*. The at-risk profiles of the present study were in line with the low-average performers and the double-deficit risk profile of the study of Ozernov-Palchik et al. (2017), as the two latter have similar profiles with different scores. However, we cannot distinguish to which of these two profiles of the study by Ozernov-Palchik et al. (2017) the at-risk profiles of the present study match best. On the other hand, although the *below-average performers with average IQ* performed lower on rapid naming tasks compared to the *below-average performers with below-average IQ*, we did not find clear evidence for distinct profiles of deficits such as a PA deficit or RAN deficit profile. Moreover, it remains unclear whether these results are in line with the double deficit hypothesis (DDH), as no profiles were found with average scores on phonological measures and below-average on rapid naming and vice versa. Second, the DDH states that children with both deficits are the poorest readers, but as this study comprised pre-reading children and includes no longitudinal data, this cannot be derived from this study.

Although a recent study found that DDH subtypes were consistently found across orthographies (Furnes, Elwér, Samuelsson, Olson, & Byrne, 2019), most evidence has been derived from English samples whereas many studies in other orthographies could not find reliable differences between the deficit groups (Torppa, Georgiou, Salmi, Eklund, & Lyytinen, 2012; Vaessen, Gerretsen, & Blomert, 2009). Nevertheless, this study demonstrated that employing an LPA in an English-speaking sample in which approximately 30% was considered as reader, resulted in partly different profiles compared to an LPA in a Dutch-speaking sample in which none of the children was considered as reader. Specifically, in our study two similar at-risk profiles emerged, whereas the study of Ozernov-Palchik et al. (2017) found evidence for three at-risk profiles.

These differences can be partly due to differences in orthographic depth, as Dutch is more consistent in its grapheme-phoneme correspondence compared to English, which could lead to differences in the relative importance of these pre-reading measures in predicting profile membership. Although PA and LK seem to be as predictive for class membership as in the English-speaking sample, the amount of variance of RAN explained by the model in this study was lower compared to the explained variance of RAN in the study of Ozernov-Palchik et al. (2017) (i.e., 0.20 compared to 0.45 respectively). This is in line with a meta-analysis of Araujo and colleagues (2015), that showed that RAN significantly predicts reading fluency and accuracy, but that this relation is moderated by orthographic depth, with weaker correlations for transparent (e.g. Dutch) compared to opaque orthographies (e.g. English). Differences in profiles could also be explained by the variability in reading experience. Vaessen and Blomert (2010) reported that the relationship between RAN and reading skills gradually increases as a function of reading experience, making RAN especially important in predicting reading speed from first grade on and later (Boets et al., 2010). As approximately 30% of the children in the study of Ozernov-Palchik et al. (2017) were considered as readers, differences in profiles could be influenced by this difference in reading experience. All members of the double-deficit risk group were non-readers, whereas only 58.3% of the RAN risk group were non-readers. Therefore, higher scores on pre-reading measures in the RAN risk group could be influenced by the performance of readers instead of the heterogeneity that would be present when all children were non-readers. Although RAN had the lowest predictive value for individual class membership in our study, we thus cannot argue that RAN is not important in predicting reading performance in Dutch-speaking readers, as our sample only comprised non-readers. Last, differences in profiles could be influenced by environmental factors. For example, English-speaking countries generally provide more reading-related activities at home and in preschool which give children a head start in mastering reading skills even prior to formal reading instruction (Mann & Wimmer, 2002).

According to the intergenerational multiple deficit model (van Bergen et al., 2014), parental and environmental influences on the children's membership classification were assessed. First, rather than dichotomizing children into high and low family risk for dyslexia, parental reading skills were included as a continuous trait. This revealed that scores on ARHQ from both parents differed significantly across profiles. Specifically, maternal and paternal ARHQ in *below-average performers* (with both average and below-average IQ) was higher than in *high performers* (with lower scores indicating less difficulty). In addition, *below-average performers* had, on average, a father with a higher ARHQ score (i.e. poorer reading ability) compared to the *average performers*. These results suggest that pre-reading skills are already influenced by parental characteristics and thus might partly determine later reading performance. This is in line with the study of van Bergen, Bishop, van Zuijen and de Jong (2015), which showed that 21% of variance in children's reading performance can be explained by parent's reading skills. Second, besides genetic influences, parents largely shape their offspring's environment. SES is traditionally defined by a grouping of financial, educational and occupational characteristics of parents (Bornstein & Bradley, 2003). However, parental education is found to be a more stable indicator of SES than family income or parental occupation (Lewis & Mayes, 2012). Moreover, parents are often less prone to provide income data (Bornstein & Bradley, 2003). Therefore, parental educational level was chosen as measure of socio-economic status in this study. This revealed that *below-average performers with below-average IQ* were more likely to have parents with a low SES whereas *high performers* were more likely to have parents with a high SES. It is possible that highly educated parents are more likely to offer a stronger home literacy environment (HLE) which stimulates the language development of the child, leading to higher scores on pre-reading measures. In contrast, parents with reading difficulties may be more likely to stimulate their children's reading development by providing a rich HLE in order to prevent reading problems (Griffin & Morrison, 1997), although some studies showed that the relationship between HLE and pre-reading skills was mainly driven by the number of books that was available at home (van Bergen, de Jong, Maassen, & van der Leij, 2014; van Bergen, van Zuijen, Bishop, & de Jong, 2017). Although these factors appeared to be predictive for children's membership classification, one should take into account the multidimensional structure of dyslexia. More specifically, according to the multiple deficit model, a variety of risk and protective factors at different levels increases or decreases the risk that a child develops a reading disorder. Therefore, although parental and environmental factors offer important insights in estimating the risk that one develops dyslexia, current results should be interpreted with caution as the relation between parental and environmental factors and the risk to develop dyslexia is likely to be moderated by other variables.

Although the LPA technique in general has some important advantages, some considerations have to be taken into account before implementing this technique in clinical practice and research. First, as there is no 'gold standard' in LPA, studies tend to use different criteria making it even more challenging to compare results of different studies. Especially when future studies show consistent results according to classification and prediction of at-risk children, general criteria should be developed, to ensure the validity of LPA in clinical practice. Second, the obtained profiles are determined by a certain test battery in a specific sample, defined by the investigators. Therefore, the emerged profiles are likely to be different when using other test batteries. Cross-validation, where one can estimate how accurately the estimated profiles can predict profiles in an independent sample, is a technique that has been proposed to validate the out-sample prediction of such a model (Arlot & Celisse, 2010). Although the strongest cognitive predictors of future reading ability have been repeatedly investigated, more research is needed to identify the optimal test battery such that results can be replicated in other studies using different samples. In addition, given the multifactorial etiology of dyslexia (Pennington, 2006), the model may lack some information to accurately reflect reality and therefore one may consider including more variables. However, combinations that constitute a small proportion are difficult to detect, making LPA especially useful in identifying subgroups that are widely represented within the general sample. Nevertheless, these limitations do not imply that the use of latent classes could not contribute to the enhancement of early intervention, but demonstrate the need for future research in order to develop complex models in which these limitations can be addressed.

Although this study carries some important implications for earlier identification and individualized interventions, it also comes with certain limitations. First, given the high classification error of 18 %, around one fifth of the children is estimated to be misclassified. This classification error shows up due to the model assignment rule, which means that children were classified to the cluster for which the posterior probability was the highest. It is therefore possible that a child was assigned to the *below-average performers* (for example because of a 52% probability), whereas it was almost equally likely that this child belonged to the *average* *performers* (with a 48% probability for example). This resulted in a large proportion of children that was classified in one of the below-average profiles which are presumably more likely to develop dyslexia. It should be acknowledged that we aimed to identify children with an elevated risk for dyslexia, and not those with a manifestation of dyslexia. Providing an intervention for all below-average performers may therefore be an overestimation of the need for intervention. However, one can consider using these probabilities instead of the modal assignment rule such that children are not required to fit into only one cluster. This way, the classification error decreases, and one can ensure that a child almost certainly belongs to an at-risk group before allocating intervention resources. Second, although this study addressed the heterogeneity within the pre-reading population, investigating longitudinal stability of class membership is important in future studies to determine which children truly develop dyslexia. This way, one can determine whether below-average performers are more likely to develop dyslexia than average and high performers. Referring to the 'dyslexia paradox', this technique allows researchers to identify which profiles are at-risk to develop dyslexia and to determine the probability that an individual has scores that are consistent with dyslexia or not, enhancing earlier interventions

In sum, the current study has shown that heterogeneity in pre-reading measures is already present prior to formal reading instruction. Specifically, two at-risk profiles emerged, which are presumably both more likely to develop dyslexia compared to the typically developing population and which might need different interventions programs. Therefore, the use of these profiles could be the foundation for early targeted intervention of dyslexia. However, more extended research is needed to identify the stability of these profiles across time and across different languages.

References

Araujo, S., Reis, A., Petersson, K. M., & Faisca, L. (2015). Rapid automatized naming and reading performance: A meta-analysis. *Journal of Educational Psychology*, *107*(3), 868–883. https://doi.org/10.1037/edu0000006

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics surveys, 4*, 40-79. https://doi.org/10.1214/09-SS054

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Boets, B., Smedt, B. de, Cleuren, L., Vandewalle, E., Wouters, J., & Ghesquière, P. (2010). Towards a further characterization of phonological and literacy problems in Dutch-speaking children with dyslexia. *British Journal of Developmental Psychology*, *28*(1), 5–31. https://doi.org/10.1348/026151010X485223

Boets, B., Wouters, J., van Wieringen, A., De Smedt, B., & Ghesquière, P. (2008). Modelling relations between sensory processing, speech perception, orthographic and phonological ability, and literacy achievement. *Brain and Language*, *106*(1), 29–40. https://doi.org/10.1016/j.bandl.2007.12.004

Bornstein, M. H., & Bradley, R. H. (Eds.) (2003). *Socioeconomic status, parenting, and child development.* Mahwah, NJ: Lawrence Erlbaum Associates.

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., … Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological Science*, *23*(6), 678–686. https://doi.org/10.1177/0956797611434536

Catts, H. W., Nielsen, D. C., Bridges, M. S., Liu, Y. S., & Bontempo, D. E. (2015). Early identification of reading disabilities within an RTI framework. *Journal of Learning Disabilities*, *48*(3), 281–297. https://doi.org/10.1177/0022219413498115

Dunn, K. M., Campbell, P., & Jordan, K. P. (2013). Long-term trajectories of back pain: Cohort study with 7-year follow-up. *BMJ Open*, *3*(12), e003838. https://doi.org/10.1136/bmjopen-2013-003838

Ferrer, E., Shaywitz, B. A., Holahan, J. M., Marchione, K. E., Michaels, R., & Shaywitz, S. E. (2015). Achievement gap in reading is present as early as first grade and persists through adolescence. *The Journal of Pediatrics*, *167*(5), 1121-1125. https://doi.org/10.1016/j.jpeds.2015.07.045

Furnes, B., Elwér, Å., Samuelsson, S., Olson, R. K., & Byrne, B. (2019). Investigating the double-deficit hypothesis in more and less transparent orthographies: A longitudinal study from preschool to grade 2. *Scientific Studies of Reading*, *23*(6), 478–493. https://doi.org/10.1080/10888438.2019.1610410

Gellert, A. S., & Elbro, C. (2017). Does a dynamic test of phonological awareness predict early reading difficulties? A longitudinal study from kindergarten through grade 1. *Journal of Learning Disabilities*, *50*(3), 227-237. https://doi.org/10.1177/0022219415609185

Georgiou, G. K., Torppa, M., Manolitsis, G., Lyytinen, H., & Parrila, R. (2012). Longitudinal predictors of reading and spelling across languages varying in orthographic consistency. *Reading and Writing*, *25*(2), 321–346. https://doi.org/10.1007/s11145-010-9271-x

Gersons-Wolfensberger, D. C., & Ruijssenaars, W. A. (1997). Definition and treatment of dyslexia: A report by the Committee on Dyslexia of the Health Council of The Netherlands. *Journal of Learning Disabilities*, *30*(2), 209–213. https://doi.org/10.1177/002221949703000208

Geurts, L., Vanden Abeele, V., Celis, V., Husson, J., Van den Audenaeren, L., Loyez, L., …

Ghesquière, P. (2015). DIESEL-X: A game-based tool for early risk detection of dyslexia in preschoolers. In J. Torbeyns, E. Lehtinen, & J. Elen (Eds.), *Describing and Studying Domain-Specific Serious Games* (pp. 93–114). https://doi.org/10.1007/978-3-319-20276-1_7

Griffin, E. A., & Morrison∗, F. J. (1997). The unique contribution of home literacy environment to differences in early literacy skills. *Early Child Development and Care*, *127*(1), 233–243. https://doi.org/10.1080/0300443971270119

Kornilov, S. A., & Grigorenko, E. L. (2018). What reading disability? Evidence for multiple latent profiles of struggling readers in a large Russian sibpair sample with at least one sibling at risk for reading difficulties. *Journal of Learning Disabilities*, *51*(5), 434–443. https://doi.org/10.1177/0022219417718833

Lefly, D. L., & Pennington, B. F. (2000). Reliability and validity of the Adult Reading History Questionnaire. *Journal of Learning Disabilities*, *33*(3), 286–296. https://doi.org/10.1177/002221940003300306

Lewis, M., & Mayes, L. C. (2012). *The Cambridge Handbook of Environment in Human Development.* Cambridge University Press.

Mann, V., & Wimmer, H. (2002). Phoneme awareness and pathways into literacy: A comparison of German and American children. *Reading and Writing, 15*(7–8), 653-682. https://doi.org/10.1023/A:1020984704781

Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, *138*(2), 322–352. https://doi.org/10.1037/a0026744

Mugnaini, D., Lassi, S., La Malfa, G., & Albertini, G. (2009). Internalizing correlates of dyslexia. *World Journal of Pediatrics*, *5*(4), 255–264. https://doi.org/10.1007/s12519-009-0049-7

Oberski, D.L. (2016). Beyond the number of classes: Separating substantive from non-substantive dependence in latent class analysis. *Advances in Data Analysis and Classification*, *10*(2), 171-182. https://doi.org/10.1007/s11634-015-0211-0

Osman, B. B. (2000). Learning disabilities and the risk of psychiatric disorders in children and adolescents. In L. L. Greenhill (Ed.), *Learning Disabilities: Implications for Psychiatric Treatment* (pp. 33–57). Washington, DC: American Psychiatric Press.

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. https://doi.org/10.1002/wcs.1383

Ozernov-Palchik, O., Norton, E. S., Sideridis, G., Beach, S. D., Wolf, M., Gabrieli, J. D. E., & Gaab, N. (2017). Longitudinal stability of pre-reading skill profiles of kindergarten children: Implications for early screening and theories of reading. *Developmental Science*, *20*(5), e12471. https://doi.org/10.1111/desc.12471

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

Pennington, B. F., Santerre-Lemmon, L., Rosenberg, J., MacDonald, B., Boada, R., Friend, A., … Olson, R. K. (2012). Individual prediction of dyslexia by single versus multiple deficit models. *Journal of Abnormal Psychology*, *121*(1), 212–224. https://doi.org/10.1037/a0025823

Peterson, R. L., & Pennington, B. F. (2012). Developmental dyslexia. *Lancet*, *379*(9830), 1997–2007. https://doi.org/10.1016/S0140-6736(12)60198-6

Poulsen, M., Nielsen, A.-M. V., Juul, H., & Elbro, C. (2017). Early identification of reading difficulties: A screening strategy that adjusts the sensitivity to the level of prediction accuracy. *Dyslexia*, *23*(3), 251–267. https://doi.org/10.1002/dys.1560

Raven, J. C., Court, J. H., & Raven, J. (1984). *Manual for Raven's progressive matrices and* *vocabulary scales*. London: Lewis.

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. https://doi.org/10.1037/0022-0663.96.2.265

Schmalz, X., Marinus, E., Coltheart, M., & Castles, A. (2015). Getting to the bottom of orthographic depth. *Psychonomic Bulletin & Review*, *22*(6), 1614–1629. https://doi.org/10.3758/s13423-015-0835-2

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174. https://doi.org/10.1348/000712603321661859

Share, D. L. (2008). On the anglocentricities of current reading research and practice: The perils of overreliance on an “Outlier” orthography. *Psychological Bulletin*, *134*(4), 584-615. https://doi.org/10.1037/0033-2909.134.4.584

Snowling, M., Bishop, D. V. M., & Stothard, S. E. (2000). Is preschool language impairment a risk factor for dyslexia in adolescence? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*(5), 587–600. https://doi.org/10.1017/S0021963099005752

Stanovich, K. E. (1996). Toward a more inclusive definition of dyslexia. *Dyslexia*, *2*(3), 154–166. https://doi.org/10.1002/(SICI)1099-0909(199611)2:3%3c154::AID-DYS63%3e3.0.CO;2-B

Torppa, M., Georgiou, G., Salmi, P., Eklund, K., & Lyytinen, H. (2012). Examining the double-deficit hypothesis in an orthographically consistent language. *Scientific Studies of Reading*, *16*(4), 287–315. https://doi.org/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*(1), 3–32. https://doi.org/10.1007/s11881-007-0003-0

Vaessen, A., & Blomert, L. (2010). Long-term cognitive dynamics of fluent reading development. *Journal of Experimental Child Psychology*, *105*(3), 213–231. https://doi.org/10.1016/j.jecp.2009.11.005

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. https://doi.org/10.1016/j.jecp.2008.12.004

van Bergen, E., de Jong, P. F., Maassen, B., & van der Leij, A. (2014). The effect of parents' literacy skills and children's preliteracy skills on the risk of dyslexia. *Journal of Abnormal Child Psychology*, *42*(7), 1187–1200. https://doi.org/10.1007/s10802-014-9858-9

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. https://doi.org/10.3389/fnhum.2014.00346

van Bergen, E., Bishop, D., van Zuijen, T., & de Jong, P. F. (2015). How does parental reading influence children's reading? A study of cognitive mediation. *Scientific Studies of Reading*, *19*(5), 325–339. https://doi.org/10.1080/10888438.2015.1050103

van Bergen, E., van Zuijen, T., Bishop, D., & de Jong, P. F. (2017). Why are home literacy environment and children's reading skills associated? What parental skills reveal. *Reading Research Quarterly*, *52*(2), 147–160. https://doi.org/10.1002/rrq.160

van den Bos, K. P., Zijlstra, B. J. H., & Spelberg, H. C. lutje. (2002). Life-span data on continuous-naming speeds of numbers, letters, colors, and pictured objects, and word-reading speed. *Scientific Studies of Reading*, *6*(1), 25–49. https://doi.org/10.1207/S1532799XSSR0601_02

Vermunt, J. K., & Magidson, J. (2005). *Latent Gold 4.0 User's Guide*. Belmont, MA: Statistical Innovations Inc.

Vermunt, J. K., & Magidson, J. (2016). *Technical Guide for Latent GOLD 5.1: Basic, Advanced, and Syntax*. Belmont, MA: Statistical Innovations Inc.

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

Willcutt, E. G., & Pennington, B. F. (2000). Psychiatric comorbidity in children and adolescents with reading disability. *Journal of Child Psychology and Psychiatry*, *41*(8), 1039–1048. https://doi.org/10.1111/1469-7610.00691

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology; Washington*, *91*(3), 415–438. https://doi.org/0022-0663/99/$3.00

Wolf, M., O'Rourke, A. G., Lovett, M., Cirino, P., & Morris, R. (2002). The second deficit: An investigation of the independence of phonological and naming-speed deficits in developmental dyslexia. *Reading and Writing: An Interdisciplinary Journal*, *15*, 43–72. https://doi.org/10.1023/A:1013816320290

Wolff, U. (2010). Subgrouping of readers based on performance measures: A latent profile analysis. *Reading and Writing*, *23*(2), 209–238. https://doi.org/10.1007/s11145-008-9160-8