

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Psyridou, Maria; Tolvanen, Asko; Patel, Priyanka; Khanolainen, Daria; Lerkkanen, Marja-Kristiina; Poikkeus, Anna-Maija; Torppa, Minna

**Title:** Reading Difficulties Identification : A Comparison of Neural Networks, Linear, and Mixture Models

**Year:** 2023

**Version:** Published version

**Copyright:** © 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.

**Rights:** CC BY 4.0

**Rights url:** <https://creativecommons.org/licenses/by/4.0/>

**Please cite the original version:**

Psyridou, M., Tolvanen, A., Patel, P., Khanolainen, D., Lerkkanen, M.-K., Poikkeus, A.-M., & Torppa, M. (2023). Reading Difficulties Identification : A Comparison of Neural Networks, Linear, and Mixture Models. *Scientific Studies of Reading*, 27(1), 39-66.

<https://doi.org/10.1080/10888438.2022.2095281>

## Reading Difficulties Identification: A Comparison of Neural Networks, Linear, and Mixture Models

Maria Psyridou, Asko Tolvanen, Priyanka Patel, Daria Khanolainen, Marja-Kristiina Lerkkanen, Anna-Maija Poikkeus & Minna Torppa

To cite this article: Maria Psyridou, Asko Tolvanen, Priyanka Patel, Daria Khanolainen, Marja-Kristiina Lerkkanen, Anna-Maija Poikkeus & Minna Torppa (2022): Reading Difficulties Identification: A Comparison of Neural Networks, Linear, and Mixture Models, Scientific Studies of Reading, DOI: [10.1080/10888438.2022.2095281](https://doi.org/10.1080/10888438.2022.2095281)

To link to this article: <https://doi.org/10.1080/10888438.2022.2095281>



© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 18 Jul 2022.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

## Reading Difficulties Identification: A Comparison of Neural Networks, Linear, and Mixture Models

Maria Psyridou <sup>a</sup>, Asko Tolvanen <sup>b</sup>, Priyanka Patel <sup>c</sup>, Daria Khanolainen <sup>c</sup>,  
Marja-Kristiina Lerkkanen <sup>c</sup>, Anna-Maija Poikkeus <sup>c</sup>, and Minna Torppa <sup>c</sup>

<sup>a</sup>Department of Psychology, Faculty of Education and Psychology, University of Jyväskylä, Jyväskylä, Finland;

<sup>b</sup>Methodology Centre for Human Sciences, Faculty of Education and Psychology, University of Jyväskylä, Jyväskylä, Finland; <sup>c</sup>Department of Teacher Education, Faculty of Education and Psychology, University of Jyväskylä, Jyväskylä, Finland

### ABSTRACT

**Purpose:** We aim to identify the most accurate model for predicting adolescent (Grade 9) reading difficulties (RD) in reading fluency and reading comprehension using 17 kindergarten-age variables. Three models (neural networks, linear, and mixture) were compared based on their accuracy in predicting RD. We also examined whether the same or a different set of kindergarten-age factors emerge as the strongest predictors of reading fluency and comprehension difficulties across the models.

**Method:** RD were identified in a Finnish sample ( $N \approx 2,000$ ) based on Grade 9 difficulties in reading fluency and reading comprehension. The predictors assessed in kindergarten included gender, parental factors (e.g., parental RD, education level), cognitive skills (e.g., phonological awareness, RAN), home literacy environment, and task-avoidant behavior.

**Results:** The results suggested that the neural networks model is the most accurate method, as compared to the linear and mixture models or their combination, for the early prediction of adolescent reading fluency and reading comprehension difficulties. The three models elicited rather similar results regarding the predictors, highlighting the importance of RAN, letter knowledge, vocabulary, reading words, number counting, gender, and maternal education.

**Conclusion:** The results suggest that neural networks have strong promise in the field of reading research for the early identification of RD.

Predicting reading difficulties (RD) accurately and early facilitates timely support. The absence of effective prediction mechanisms can result in delayed RD identification, putting children at increased risk of lower academic achievement and of developing low self-concept, anxiety, and even depression (e.g., Aro et al., 2019; McArthur et al., 2021; Sanfilippo et al., 2020). Despite advances in the early identification and prediction of RD, many children who would benefit from early enrichment and support remain unidentified and without timely access to preventive environments. One key reason why early identification has proved to be challenging is that there is no single deficit at the core of RD, but rather a combination of various deficits that can vary from child to child (e.g., Pennington, 2006; Van Bergen et al., 2014). In response to this, hybrid models have been suggested as a promising approach (e.g., Erbeli et al., 2018; Schatschneider et al., 2016; Spencer et al., 2014). Hybrid model thinking entails that accurate identification of RD requires a complex multivariate assessment in which predictions are improved by using multiple predictor measures. Thus, statistical approaches that can handle a large number of predictors with differential impacts on different individuals are necessary.

**CONTACT** Maria Psyridou  maria.m.psyridou@jyu.fi 

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Predictors of RD

In this study, we focused on the prediction of RD in decoding and in reading comprehension (RC) in adolescence, which remains an understudied age group (Eklund et al., 2018; Manu et al., 2021; Ricketts et al., 2020). Since we focused on adolescent reading skills and our sample concerns a language with a highly transparent orthography (Finnish), our focal decoding measure is reading fluency (RF) and not accuracy, which by this age and in this context does not present variation that would have discriminative value (Aro, 2017).

Early cognitive skills are powerful predictors of later RD (e.g., Caravolas et al., 2019; Clayton et al., 2020; Psyridou et al., 2021). The literature indicates partially different predictors of RF and RC. Of the cognitive predictors of RD in RF, we focused on phonological awareness, letter knowledge, RAN, and number counting, which emerge as key early predictors of RF both among Finnish-speaking children (Kairaluoma et al., 2013; Koponen et al., 2016; Psyridou et al., 2021;), and in other language contexts (e.g., Bernabini et al., 2021; Caravolas et al., 2019; Clayton et al., 2020; Melby-Lervåg et al., 2012; Snowling et al., 2019). Studies on the early predictors of RC have shown strong predictive effects of early vocabulary (e.g., Florit & Cain, 2011; Manu et al., 2021; Nation et al., 2010; Psyridou et al., 2018) and listening comprehension measures (e.g., Florit & Cain, 2011; Manu et al., 2021; Nation et al., 2010). Therefore, we included vocabulary and listening comprehension among the cognitive predictors of RD in this study.

Motivational factors also are important predictors of children's learning (e.g., Schiefele et al., 2012). In the present study, motivation was represented by task-avoidant behavior which in prior studies has been shown to be more prevalent among children with poor reading skills and may further impede their reading progress (Georgiou et al., 2010; Hirvonen et al., 2010). Longitudinal research has indicated that higher task avoidance predicts slower reading development, even after controlling for motivational factors such as autoregressors or phonological processing and non-verbal IQ (Georgiou et al., 2010).

Home literacy environment, particularly shared parent-child book reading and instruction of literacy skills at home, has also been found to predict children's language and literacy development (e.g., Khanolainen et al., 2020; Mol & Bus, 2011; Silinskas et al., 2020a). Some evidence suggests, however, that home literacy environment may simply reflect parental skills and the correlation between home environment and child skills may in fact be due to the correlation between the skills of parents and their children. Puglisi et al. (2017) and Van Bergen et al. (2017) reported that the correlation between measures of children's literacy skills and home literacy environment largely lost its significance after controlling for parental literacy skills. In the light of this, research that examines the effects of home literacy environments should also include parental reading measures. Parental RD (or family risk for RD) have been identified as one of the best early predictors of children's RD (Esmaeli et al., 2019; Psyridou et al., 2021; Snowling & Melby-Lervåg, 2016; Van Bergen et al., 2015). In the present study, we included family risk for RD, at-home learning activities, as well as parental education level as predictors of children's RD. Inclusion of parental education level is in line with studies documenting strong effects of parental education levels on children's reading development (e.g., Hamilton et al., 2016). Finally, the child's gender was included as a predictor as previous studies have suggested increased vulnerability for RD among boys (e.g., Manu et al., 2021; Psyridou et al., 2021; Quinn & Wagner, 2015).

## Statistical models for the prediction of RD

There are multiple potential statistical models available for early identification of risk factors for RD. In the present study we focus on comparing two commonly used but differing approaches (linear regression and mixture model) and a newer neural network-based approach. The deep artificial neural networks, a category of deep learning and a subcategory of machine learning (Goodfellow et al., 2016; Urban & Gates, 2021) is a promising new methodology that has been used successfully in other fields,

but still is used rarely in reading research. Deep neural networks are computational models inspired by the human brain and have impacted research on natural language processing (Graves et al., 2013) and gaming (Mnih et al., 2015; Silver et al., 2016), among others. They are composed of units that work in parallel and are arranged in interconnected layers: an input layer that includes the independent variables; one or more hidden layers that contain unobservable network units; and an output layer which includes the dependent variables (Hinton, 2007; LeCun et al., 2015). A deep neural network learns to perform a specific task (e.g., prediction, or classification) through training during which the model learns the strength of the connections between the units in the layers (Cichy & Kaiser, 2019; Urban & Gates, 2021). Once trained, the deep neural networks model can be used to perform the same task using new inputs (Cichy & Kaiser, 2019).

The neural networks models have the advantage of being able to handle a large number of variables that can correlate with each other and do not need to be linearly associated (Urban & Gates, 2021). Thus, in addition to the linear effects, the neural networks models can identify linear, nonlinear, or interactional effects (for example, several combinations of the independent variables building cumulative risk) or a combination of these. In contrast to neural networks, when using linear models, only linear effects can be identified and it is often difficult to include a broad set of predictors in a linear prediction model due to collinearity problems.

Mixture models, which along with the linear regression model, is widely used in reading research, can handle a large number of correlated variables and do not expect that the associations between variables are the same for all participants of the sample. Instead, mixture models (e.g., latent class analysis, latent transition analysis, latent profile analysis) allow for the identification of homogeneous groups from a heterogeneous sample. From the prediction perspective, the identification of subgroups among predictors (e.g., identification of differential risk factor profiles for RD) can reveal combinations of predictors that are meaningfully associated with the outcome; this may not be identifiable in models which assume the same estimates across all participants. Even though mixture models were not designed for optimizing prediction per se, they provide the possibility to identify groups of children for whom the probability of belonging to the RD group is very high (even 1.0).

## The present study

The present study aims to identify the most accurate model for predicting adolescent (Grade 9) RF and RC difficulties using a large set of kindergarten-age variables. Three models (neural networks, linear, and mixture) were compared based on their accuracy to predict RD, and to determine whether the same or different set of kindergarten-age factors emerge as the strongest predictors of RF or RC difficulties in the three models. The neural networks model was selected because it is a new method which has been broadly used in other fields with very promising results. We sought to examine whether the neural networks model can provide as high, or higher, accuracy in the identification of RD than the more commonly used methods. The linear model and mixture model were selected as contrast models to the new neural networks model because they are commonly used, have a differential approach to prediction but both use maximum likelihood methods, and there is a robust evidence of their usefulness in reading research.

To the best of our knowledge, no previous studies have compared these three models in the field of reading research. Previous studies in other fields that have used neural networks for prediction (without comparing them with linear and/or mixture models) have demonstrated high accuracy (e.g., Alzheimer's: Lu et al., 2018; ADHD: Kuang & He, 2014; autism: Heinsfeld et al., 2018; biomedicine: Mamoshina et al., 2016; Parkinson's: Choi et al., 2017; psychiatry: Vieira et al., 2017). Given these results, we hypothesized that neural networks would provide high accuracy in predicting RD.

## Method

### Participants

The present study was part of the Finnish longitudinal First Steps Study that includes data from approximately 2,000 children from kindergarten to Grade 9 (Lerkkanen et al., 2006). Reading skills of the participants were assessed in Grades 1, 2, 3, 4, 6, 7, and 9. In the present study, we used data from children's assessment taken in the fall and spring of kindergarten (i.e., age 6), and their RF and RC skills assessment taken in Grade 9. There were 1,880 children at the beginning of the follow-up but when they entered school, all their classmates were also invited to participate. Over the years the sample size varied and altogether 2,518 children participated in at least one assessment. Descriptive statistics for the sample used in this study can be found in Table 2 (more information on the number of participants in each assessment and the descriptive statistics for the reading measures used in Grades 1–9 can be found in Psyridou et al., 2021). The sample was drawn from four municipalities – two in central, one in western, and one in eastern Finland – out of which one was mainly urban, one mainly rural, and two included both urban and semi-rural environments. In three of the municipalities, the participants comprised the entire age cohort of children, and in another municipality, the participating children comprised about half the age cohort. Out of the parents who were contacted, 78–89% agreed to participate in the study. Ethnically and culturally, the sample was very homogeneous and representative of the Finnish population, and parental education levels were very close to the national distribution of Finland (Statistics Finland, 2007). The university's Ethical Committee approved the First Steps Study, and the study conforms to the Declaration of Helsinki. All participants provided informed written consent and children gave their assent prior to their inclusion in the study.

### Measures

The children were assessed longitudinally, in kindergarten (fall 2006 and spring 2007) and in Grade 9 (RF: spring 2016; RC (PISA): fall 2015). The measures are described in detail in Table 1.

### Statistical analysis

Kindergarten-age measures were used to predict RF and RC difficulties in Grade 9 using three models (neural networks, linear, and mixture). RF and RC were analyzed separately, and RF or RC difficulty was defined as scoring in the lowest 10% of the RF or RC distribution, respectively.

### Neural networks model

For the neural networks model, the Multilayer Perceptron Network (MLP) was used to produce a predictive model to identify RF or RC difficulties based on the kindergarten-age measures. MLP was conducted using SPSS (version 26). We first examined whether the missing-completely-at-random (MCAR) assumption could be established for the data. Little's MCAR test indicated that the data were not MCAR (Little, 1988). The effect sizes (Cohen's  $d$ ) for the comparison of the participants with and without data ranged from .01 to  $-.39$ . However, because it is important for MLP not to have missing values, an expectation-maximization algorithm in SPSS was used to impute missing data using only the kindergarten-age variables so that the Grade 9 reading variables remained unchanged and independent from the kindergarten-age variables.

The MLP was set to randomly choose approximately 70% of the data for the training model and 30% for the testing model. Furthermore, we allowed the model to choose the number of hidden layers and units automatically. The training sample was used for the estimation of the weights. Once the weights had been decided, the testing sample was utilized to estimate whether the weights generated with the training sample apply to another subset of people and whether the model can be generalized. To balance the imputed cases between the training and testing models, the selection of participants for both models is important. Therefore, we used 20 different seed values to select and allocate

**Table 1.** Reading measures and kindergarten-age factors used in the study.

Measure	Task	Assessment Year	Description	Scoring	Reliability
Reading Fluency	Word reading fluency (ALLU–Ala-asteen lukutesti [ALLU–Reading Test for Primary School]; Lindeman, 2000)	Grade 9, Spring 2016	80 items consisting of a picture of a common item presented along with four orthographically and phonologically similar words (e.g., pupu (bunny), pipo (cap), papu (bean), apu (help) (English words are in parentheses)). Students have 2 minutes to silently match the picture to the correct word.	Number of correctly matched pictures and words. Max 80.	Cronbach's alpha of reading fluency mean score = .84
	Word-chain task (Nevala & Lyytinen, 2000)		10 word chains consisting of 4–6 words written without spaces. Students have 90 seconds to silently read and mark the word boundaries.	Number of correctly marked divisions. Max 40.	
	Sentence reading (YKÄ; Lerkkänen et al., 2018)		Standardized test consisting of 70 sentences. Students are given 2 minutes to silently read and verify the truthfulness of the sentences.	Number of correctly marked sentences. Max 70.	
Reading Comprehension	PISA Reading link items (OECD, 2010)	Grade 9, Fall 2015	Students were given 60 minutes to read 8 different sets of reading material and answer 31 questions. The reading material consisted of texts, tables, graphs, and figures. There were 15 multiple-choice questions and 16 questions which required written responses of which 12 required accessing and retrieving information, 12 required integrating and interpreting information, and 7 required reflecting and evaluating information.	A score of 1 was given for every correct response. Max 31.	Cronbach's alpha = .75
Kindergarten Measures	Initial phoneme identification (ARMI; Lerkkänen et al., 2006)	Fall 2006, Spring 2007	10 sets of 4 pictures, each depicting an object. Students were first asked to name aloud the objects and then identify the object with the same initial phoneme as the one spoken aloud by the assessor. All sounds were single phonemes.	A score of 1 was given for every correctly selected object. Max 10.	Cronbach's alpha = .75 (fall), .71 (spring)
	Letter knowledge (ARMI; Lerkkänen et al., 2006)	Fall 2006, Spring 2007	29 uppercase letters arranged in random order across three rows. Students were shown the letters on a sheet row-by-row and asked to name them aloud. Either a phoneme or letter name was regarded as correct. The test was discontinued after 6 incorrect responses.	A score of 1 was given for every correct response. Max 29.	Cronbach's alpha = .94 (fall), .93 (spring)
	Receptive Vocabulary (PPVT-R, Form L; Dunn & Dunn, 1981)	Spring 2007	30-item version of the Peabody Picture Vocabulary Test-Revised. Students were required to select the picture, out of 4 options, that correctly depicts a spoken word.	A score of 1 was given for every correct response. Max 30.	Cronbach's alpha = .60
	RAN (Denckla & Rudel, 1976)	Spring 2007	The children were asked to name as fast as possible a series of five pictures of objects arranged in semi random order in five rows of 10. There was a practice trial before the test to ensure the child's familiarity with names of the objects.	Total matrix completion time in seconds.	-

*(Continued)*

Table 1. (Continued).

Measure	Task	Assessment Year	Description	Scoring	Reliability
	Reading Words (ARMI; Lerikkanen et al., 2006)	Fall 2006, Spring 2007	Students were administered a word list containing 6 words at the fall assessment and 10 words at the spring assessment. Students were asked to read aloud the words. At the fall assessment, there were 4 two-syllabic words, 1 three-syllabic word and 1 five-syllabic word. At the spring assessment, there were 7 two-syllabic words, 2 three-syllabic words, and 1 five-syllabic word.	A score of 1 was given for every correctly read word. Max 10.	Cronbach's alpha = .84 (fall), .85 (spring)
	Listening comprehension (Vauras et al., 1995)	Spring 2007	Groups of 6 students were read aloud a story (130 words), twice, and then asked six multiple-choice questions based on the story, one question at a time. In four of the questions there were three choices, and in two questions there were four choices. Each question was accompanied by 3 or 4 pictures and student responded by marking the picture that correctly matched the story in their own test booklet.	2 points were given for every correct answer. Max 12.	Cronbach's alpha = .30/ Revelle's omega = .42
	Number counting (for similar tasks, see, Koponen et al., 2007)	Fall 2006, Spring 2007	Pre-math skills were assessed through four tasks in which children were asked to count aloud forward (from 1 to 31 and from 6 to 13) and backward (from 12 to 7 and from 23 to 1).	Scored using a 3-point scale: 2 = no errors, 1 = one small error, 0 = two or more errors. Max 8.	Cronbach's alpha = .48 (fall), .62 (spring)/ Revelle's omega = .52 (fall), .87 (spring)
	Task-avoidant behavior (Behavior Strategy Rating Scale; Zhang et al., 2011)		Kindergarten teachers evaluated the behavior of each student in the class by rating them on 5 questions based on how the child typically behaved in classroom situations (e.g., Does the child have a tendency to find something else to do instead of focusing on the task at hand?; Does the child show persistence even in the more difficult tasks?). Ratings were done on a 5-point Likert scale (1 = not at all; 5 = to a great extent).	A sum score of the 5 items was calculated.	Cronbach's alpha = .92

(Continued)



Table 1. (Continued).

Measure	Task	Assessment Year	Description	Scoring	Reliability
Parental Questionnaire Measures	Maternal education	Spring 2007	Mothers were asked to indicate their own education level on a 7-point scale: 1 = no vocational education, 2 = vocational courses (4 months), 3 = vocational school degree, 4 = vocational college degree, 5 = polytechnic degree or bachelor's degree, 6 = master's degree, and 7 = licentiate or doctoral degree.	Answers were recoded using a 3-point scale: basic education, vocational education, and university education.	-
	Home learning environment (Sénéchal et al., 1998; see, Silinskas et al., 2020b)		3-item questionnaire about at-home learning activities answered by the mothers. It included 1-item regarding shared reading which was answered using a 5-point Likert scale (1 = less than once a week, 2 = 1–3 times a week, 3 = 4–6 times a week, 4 = once a day, 5 = more than once a day) and 2-items regarding in-home teaching of literacy (teaching letters & teaching reading) which was also answered in a 5-point scale (1 = never at all/rarely to 5 = very often/daily).	Each item was examined individually.	-
	Family RD		Parents were asked to indicate on a 3-point scale whether they had clear difficulties, some difficulties, or no difficulties.	A child was considered as having family risk if the mother or the father reported that she or he had experienced some or clear RD	-

participants to the training and testing models, randomly, using the continuous RF and RC variables. Because MLP results vary from analysis to analysis, the MLP was run 20 times using the same seed and the predicted RF and RC scores from each analysis were saved (i.e., 20 seeds were run 20 times each, separately for RF and RC). Next, the mean of the 20 predicted RF/RC scores was calculated for each seed and the correlations between the mean of the predicted RF/RC and the observed RF/RC scores for the training and testing model were estimated for each seed. For the 20 seeds, the correlations between the mean of the 20 predicted RF/RC scores and the observed RF/RC scores in the training (70% of data) and testing (30% of data) models varied from .49 to .62 for the RF models ([Appendix B](#)) and from .48 to .58 for the RC models ([Appendix C](#)). The seed with the closest and highest correlation among the training and testing model was used for the prediction of RF/RC difficulties.

Next, using the best-identified seed, the MLP was run 20 times to predict the dichotomized RF and 20 times to predict the dichotomized RC variables with the kindergarten variables, and the predicted RF/RC scores from each analysis were saved. As the dependent variables are dichotomous, the predicted scores represent each participant's probability of belonging to the lowest 10% of the RF or RC distributions.

The average of the 20 predicted scores was used to test the ability of the model to predict RF and RC difficulties using Receiver Operating Characteristic (ROC) curves (Fawcett, 2006). The ROC curve is plotted with the true-positive rate (i.e., sensitivity) against the false-positive rate (i.e., 1-specificity) where the true-positive rate is on the y-axis and false-positive rate is on the x-axis. ROC curves compare sensitivity versus 1-specificity across a range of values for the ability to predict a dichotomous outcome. Each point on the ROC curve represents a sensitivity/1-specificity pair corresponding to a particular cutoff. The ROC curve is a useful tool because the curves of different models can be directly compared, either in general or for different thresholds, and the area under the curve can be used as a summary of the model skill. The larger the area under the curve, the better the identification of those with and without difficulties.

In addition to the ROC curves, we attempted to identify the most important kindergarten-age factors in predicting RF and RC difficulties. We conducted an independent variable importance analysis (provided by the MLP), which computes each kindergarten-age factor's importance in determining the neural network based on the combined training and testing samples. The analysis was conducted 20 times, and the mean of the normalized importance for each kindergarten-age factor was calculated based on the results from the 20 analyses.

### **Linear model**

For the linear model, a logistic regression analysis with backward deletion (BReg) was used to identify the kindergarten factors that predict RF or RC difficulties in Grade 9. BReg was conducted in Mplus (version 8.7). Logistic regression is useful when we aim to predict the presence or absence of a characteristic based on a set of predictors. It is similar to linear regression but is more suitable when the dependent variable is dichotomous (in this case, RD vs. no RD). Even though stepwise regression (either forward or backward) has received criticism, it was selected because it allows us to prune a list of plausible explanatory variables down to a parsimonious group of the "most important" variables. In addition, this criticism is mitigated by having a separate training and testing sample. We used the backward selection because starting with the full model has the advantage of considering the effects of all variables simultaneously. This is especially important in cases of collinearity because backward selection may be forced to keep all the variables in the model unlike forward selection where none of them may be entered. Full information maximum likelihood was used as an estimator, and robust standard errors were calculated (MLR estimator in Mplus). The MLR allows the calculation of more unbiased standard errors. BReg analysis was done in a stepwise manner. All kindergarten-age factors were added to the model, then the factor with the highest p-value was removed (one factor at a time) until all remaining factors were statistically significant. By using the full sample ( $N = 1,880$ ) without the testing sample as defined by the MLP ( $N = 1,535$  for RF,  $N = 1,576$  for RC), we identified the best minimal group of factors that predicted RF or RC difficulties. Predictive values for each

individual were saved and used for the ROC curves and comparison of the models. ROC curves were produced using the same testing sample as in the MLP (N = 345 for RF; N = 304 for RC) to examine the model's ability to predict RF or RC difficulties.

### ***Mixture model***

For the mixture model, a latent profile analysis (LPA) was used to identify homogeneous profiles for the kindergarten-age measures based on the whole data set (N = 1,880). LPA was conducted in Mplus (version 8.7). To ensure each profile's validity, we used 500 starting values, as a large set of random starting values is recommended (Asparouhov & Muthen, 2008). In the best-fitting LPA model, the individuals are assigned to a specific profile based on the highest posterior probabilities. This information was saved to an Mplus file and used further in SPSS to compare the results with the other two models. Given that for the LPA we were unable to produce ROC curves, a cross-tabulation between the LPA profile grouping and the dichotomized RF and RC variables was used.

### ***Model comparison***

Finally, using the same 30% of the data used for the testing round in MLP and estimating the linear model, the three models' results were compared using a logistic regression to examine which method can produce the specific information that the others cannot achieve. The maximum likelihood function in logistic regression gives a chi-square test which tests the fit of the model. The dichotomized RF or RC variable was used as the dependent variable and was predicted using the predicted MLP scores (i.e., the predicted probability of the individual belonging to the lowest 10% of the RF or RC distribution based on the MLP analysis), the predictive BReg scores (i.e., the predicted probability of the individual belonging to the lowest 10% of the RF or RC distribution based on the BReg analysis), and the saved latent class information from the LPA (i.e., to which profile each individual belonged). The prediction scores and latent class information of the three models were used as independent variables. One method's predictive scores were added in the first block, and the predictive scores from the other method were added in the second block. The chi-square test indicates whether the predictor in the second block is more informative for predicting the dichotomous RF or RC variable than the predictor in the first block. For the most informative method, we also examined whether it was more informative than the other two methods combined.

Computational codes for all the analyses can be found in Supplemental Materials.

## **Results**

### ***Descriptive statistics***

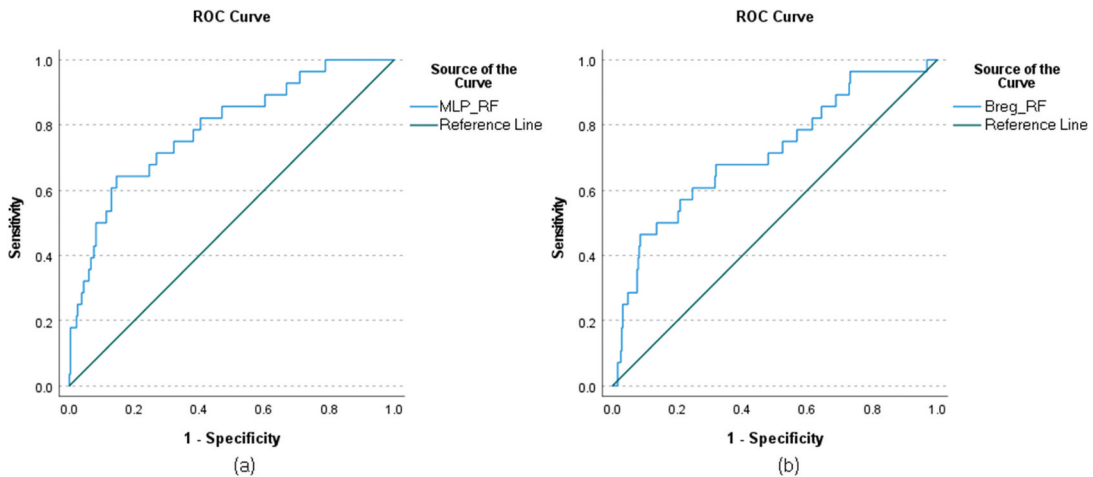
The descriptive statistics of the kindergarten-age factors and the RF and RC measures in Grade 9 are presented in Table 2. The correlations between all assessed measures are displayed in Appendix A. Both RF (mean score) and RC significantly correlated with all kindergarten-age factors, except for teaching letters at home.

### ***Neural networks model***

For the best identified seed, the correlation between the predicted and observed RF scores was .58 for the training model and .58 for the testing model, while the correlation between the predicted and observed RC scores was .55 for the training model and .55 for the testing model. Using the best identified seed, the MLP was run 20 times to predict the dichotomized RF and 20 times to predict the dichotomized RC variables with the kindergarten variables. The average of the 20 predicted scores for RF and RC represents the predicted probability that the individual belongs to the lowest 10% of the respective distribution. The average of the 20 predicted scores for RF and RC were used to produce the ROC curves for RF and RC (Figures 1a and 2a for the testing models; Appendix Z for the training models) and then

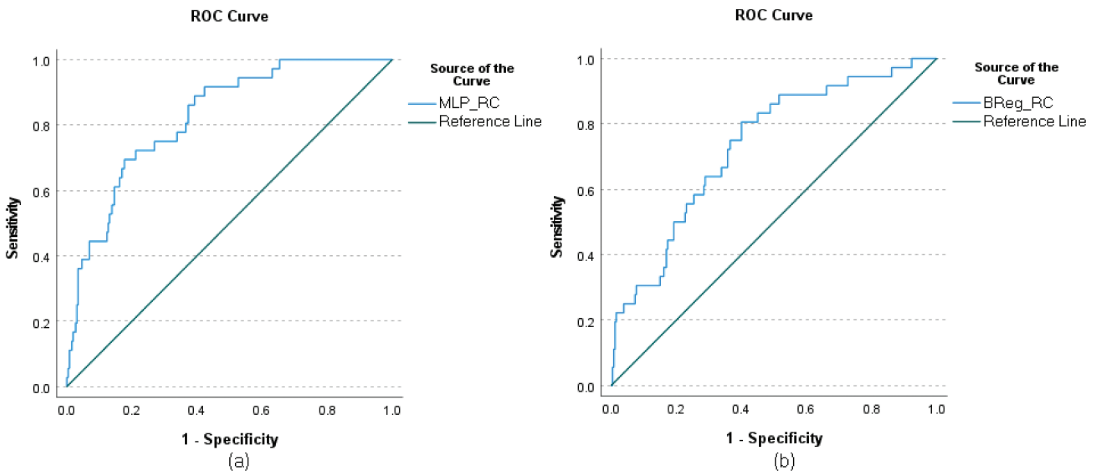
**Table 2.** Descriptive statistics for the kindergarten-age factors and the reading fluency and reading comprehension skills in Grade 9.

Measure	N	Minimum	Maximum	Mean	S.D.	Skewness	Kurtosis
Kindergarten measures							
Maternal education	1558	1.00	4.00	2.84	.90	-.14	-1.00
Family risk for RD	1115	.00	1.00	.36	.48	.59	-1.65
Teaching letters	1568	1.00	5.00	2.94	.97	.05	-.34
Teaching reading	1569	1.00	5.00	2.24	1.00	.48	-.41
Shared reading	1559	1.00	5.00	2.92	1.15	-.15	-1.01
Gender	1880	1.00	2.00	1.52	.50	-.09	-1.99
Phonological awareness fall	1867	.00	10.00	7.46	2.45	-.81	-.21
Phonological awareness spring	1836	.00	10.00	8.93	1.72	-1.99	4.11
Letter knowledge fall	1867	.00	29.00	16.95	9.01	-.25	-1.27
Letter knowledge spring	1836	.00	29.00	23.21	6.61	-1.34	1.02
Reading words fall	1867	.00	6.00	1.00	1.99	1.81	1.60
Reading words spring	1823	.00	10.00	4.03	4.29	.44	-1.61
Number counting fall	1866	.00	8.00	4.43	2.83	-.21	-1.33
Number counting spring	1836	.00	8.00	6.06	2.20	-1.10	.25
RAN	1835	.34	210.00	173.71	17.78	-1.72	6.69
Vocabulary	1839	7.00	29.00	19.82	3.38	-.38	.31
Listening comprehension	1832	.00	12.00	7.71	2.34	-.31	-.13
Task avoidant behavior	1814	1.00	5.00	3.68	1.03	-.55	-.55
Reading measures in Grade 9							
Word list reading	1707	11.00	72.00	41.47	9.07	-.02	.27
Differentiate words	1704	8.00	76.00	46.06	11.55	-.17	-.10
Sentence reading	1705	.00	70.00	37.53	8.48	-.05	.38
Reading fluency mean score	1721	-2.60	2.60	.00	.87	-.09	-.14
Reading comprehension (PISA)	1512	.00	32.73	20.26	6.20	-.56	-.04



**Figure 1.** ROC curves for reading fluency using MLP (figure a) and BReg (figure b) for the testing models. The area under the curve for the MLP is .79 and for the BReg is .72. In the testing sample there were 345 cases. ROC curves for the training models can be found in Appendix Z.

to compare the models. The ROC curves suggested that the MLP had high classification accuracy; the area under the curve was .79 ( $p < .001$ , 95% CI .699 – .880) for RF, and .83 ( $p < .001$ , 95% CI .764 – .890) for RC. Based on the ROC analysis, for RF when the sensitivity was .82 the 1-specificity was ranging from .40 to .47 and the precision was ranging from .13 to .15. For RC when the sensitivity was .82, the



**Figure 2.** ROC curves for reading comprehension using MLP (figure a) and BReg (figure b) for the testing models. The area under the curve for the MLP is .83 and for the BReg is .75. In the testing sample there were 304 cases. ROC curves for the training models can be found in Appendix Z.

1-specificity was .37 and the precision was .24. (See also, [Tables 3 and 4](#) for the corresponding sensitivity, specificity, and precision estimates for different cutoff values and [Tables 5 and 6](#) for specificity and precision values corresponding to specific sensitivity values).

Finally, the independent variable importance analysis indicated that the most important factors for predicting RF difficulties were RAN, letter knowledge (spring), and gender ([Figure 3](#)). For the prediction of RC difficulties, vocabulary, RAN, and letter knowledge (spring) emerged as the most important factors ([Figure 4](#)).

**Linear model**

The estimates and p-values for each BReg step are presented in [Appendix D](#) for RF and [Appendix E](#) for RC. In the final RF model ([Table 7](#)), four significant predictors remained: gender, letter knowledge (spring), number counting (spring), and RAN. The model explained 32% of the variance in Grade 9 RF. In the final RC model ([Table 8](#)), five significant predictors remained: maternal education level, gender, reading words (spring), RAN, and vocabulary. The model explained 39% of the variance in Grade 9 RC.

**Table 3.** Corresponding sensitivity, specificity, and precision estimates for different cutoff values for reading fluency.

cutoff	MLP					BReg				
	Sensitivity	1-Specificity	Specificity	Precision	F-measure	Sensitivity	1-Specificity	Specificity	Precision	F-measure
lowest 10%	1	.893	.107	.090	.165	.964	.893	.107	.087	.160
lowest 15%	1	.836	.164	.096	.175	.964	.842	.158	.092	.168
lowest 20%	.964	.785	.215	.098	.178	.964	.789	.211	.097	.176
lowest 25%	.964	.732	.268	.104	.188	.964	.735	.265	.104	.188

For the testing data out of the 345 cases, the true positive cases identified both by the MLP and the BReg were 28.

**Table 4.** Corresponding sensitivity, specificity, and precision estimates for different cutoff values for reading comprehension.

cutoff	MLP					BReg				
	Sensitivity	1-Specificity	Specificity	Precision	F-measure	Sensitivity	1-Specificity	Specificity	Precision	F-measure
lowest 10%	1	.887	.113	.139	.244	.974	.891	.109	.135	.237
lowest 15%	1	.831	.169	.147	.256	.947	.835	.165	.140	.244
lowest 20%	1	.774	.226	.156	.270	.947	.782	.218	.148	.256
lowest 25%	1	.718	.282	.166	.285	.921	.726	.274	.154	.264

For the testing data out of the 304 cases, the true positive cases identified both by the MLP and the BReg were 38.

**Table 5.** Specificity and precision values corresponding to specific sensitivity values for reading fluency.

MLP				BReg			
Sensitivity	1-Specificity	Specificity	Precision	Sensitivity	1-Specificity	Specificity	Precision
.893	.603-.669	.331-.397	.105-.116	.893	.688-.729	.271-.312	.098-.103
.786	.382-.404	.596-.618	.147-.154	.786	.568-.615	.385-.432	.101-.109
.714	.268-.322	.678-.732	.164-.190	.714	.479-.524	.476-.521	.108-.116

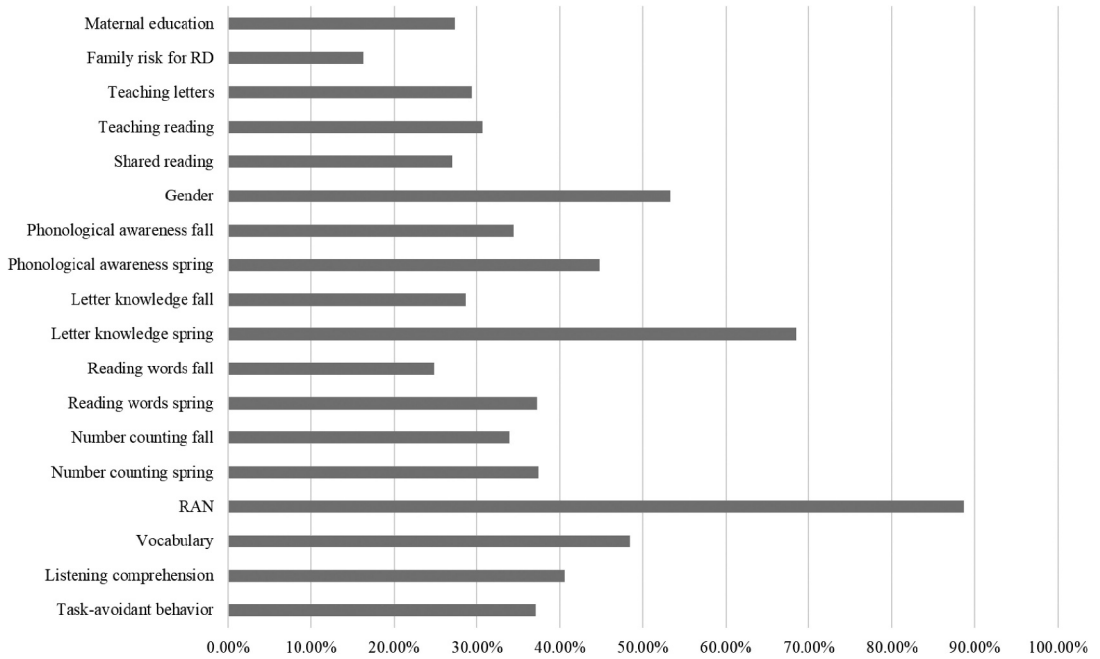
For the testing data out of the 345 cases, the true positive cases identified both by the MLP and the BReg were 28.

**Table 6.** Specificity and precision values corresponding to specific sensitivity values for reading comprehension.

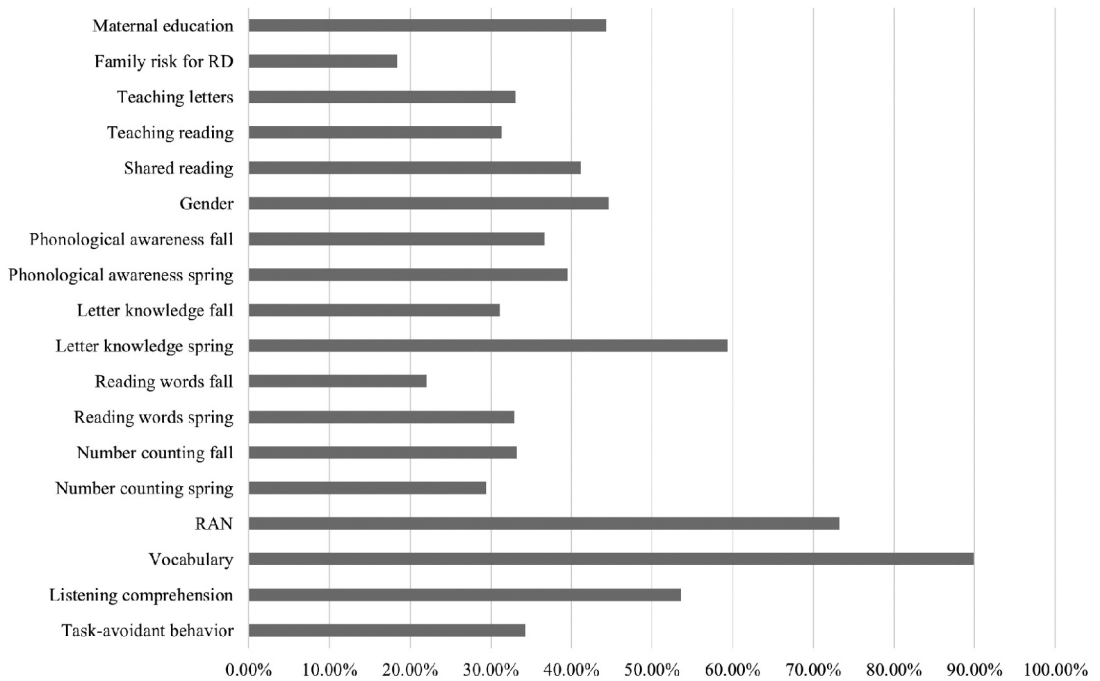
MLP				BReg			
Sensitivity	1-Specificity	Specificity	Precision	Sensitivity	1-Specificity	Specificity	Precision
.895	.395-.425	.575-.605	.231-.245	.895	.519-.662	.338-.481	.162-.198
.789	.338-.365	.635-.662	.236-.250	.789	.406	.594	.217
.711	.180-.214	.786-.820	.321-.360	.711	.357	.643	.221

For the testing data out of the 304 cases, the true positive cases identified both by the MLP and the BReg were 38.

Using the regression equation with estimated coefficients for the significant predictors of RF and RC, we calculated predictive values for each individual, representing the predicted probability that the individual belongs to the lowest 10% of the respective distribution. These predictive values first were used for the ROC curves, and then for comparing the models. Similar to the MLP, ROC curves were produced using 30% of the data to examine the model's ability to predict RF or RC difficulties (Figure 1b, 2b). The ROC curves suggested good accuracy in identifying RF and RC difficulties; the area under the curve was .72 ( $p < .001$ , 95%CI .607 – .822) for RF and .75 ( $p < .001$ , 95%CI .664 – .828) for RC. Based on the ROC analysis, for RF when the sensitivity was .82 the 1-specificity was ranging from .62 to .64 and the precision was ranging from .10 to .11. For RC, when the sensitivity was .82 the 1-specificity was ranging from .41 to .46 and the precision was ranging from .20 to .22. (See also, Tables 3 and 4 for the corresponding sensitivity, specificity, and precision estimates for different cutoff values and Tables 5 and 6 for specificity and precision values corresponding to specific sensitivity values).



**Figure 3.** Mean of normalized importance for the kindergarten-age factors for the prediction of reading fluency difficulties.



**Figure 4.** Mean of normalized importance for the kindergarten-age factors for the prediction of reading comprehension difficulties.

**Table 7.** Estimates and the p-values of the final multiple regression model with backward deletion for the prediction of reading fluency difficulties.

Measure	Final model (Step 15)	
	Estimate	p-value
Maternal education	*	*
Family risk for RD	*	*
Teaching letters	*	*
Teaching reading	*	*
Shared reading	*	*
Gender	.84	.00
Phonological awareness fall	*	*
Phonological awareness spring	*	*
Letter knowledge fall	*	*
Letter knowledge spring	-.40	.00
Reading words fall	*	*
Reading words spring	*	*
Number counting fall	*	*
Number counting spring	-.31	.02
RAN	-.36	.00
Vocabulary	*	*
Listening comprehension	*	*
Task-avoidant behavior	*	*

The standardized estimates are presented for each factor along with the p-value. Asterisks represent factors that have been deleted in previous steps. The estimates and the p-values for each BReg step are presented in Appendix D. For the identification of the best minimal group of predictors we used the full sample without the imputation of the data excluding the testing sample (N = 1,535).

**Table 8.** Estimates and the p-values of the final multiple regression model with backward deletion for the prediction of reading comprehension difficulties.

Measure	Final model (Step 14)	
	Estimate	p-value
Maternal education	-.43	.00
Family risk for RD	*	*
Teaching letters	*	*
Teaching reading	*	*
Shared reading	*	*
Gender	.59	.00
Phonological awareness fall	*	*
Phonological awareness spring	*	*
Letter knowledge fall	*	*
Letter knowledge spring	*	*
Reading words fall	*	*
Reading words spring	-.60	.00
Number counting fall	*	*
Number counting spring	*	*
RAN	-.32	.01
Vocabulary	-.58	.00
Listening comprehension	*	*
Task-avoidant behavior	*	*

The standardized estimates are presented for each factor along with the p-value. Asterisks represent factors that have been deleted in previous steps. The estimates and the p-values for each BReg step are presented in Appendix E. For the identification of the best minimal group of predictors we used the full sample without the imputation of the data excluding the testing sample (N = 1,576).



### Mixture model

Nine latent profile solutions were tested and compared, each testing a different number of profiles (1 through 9; Table 9). The model with eight profiles was considered the best-fitting model, as VLMR and LMR indicated that the models with more than eight profiles were not significantly better. The average latent class probabilities for belonging to a profile were very high, suggesting that the eight profiles were very distinct: .95 for Profile 1; .94 for Profile 2; .92 for Profile 3; .99 for Profile 4; .99 for Profile 5; .98 for Profile 6; .99 for Profile 7; and .99 for Profile 8.

As Figure 5 and Table 10 show, the eight identified profiles included two profiles with below-average scores in all kindergarten skills and six groups with average or above-average scores. Profile 1 had low scores in all cognitive skills (particularly in phonological awareness, letter knowledge, and number counting) and had family risk for RD. Profile 2 also had below-average scores in all kindergarten assessments, but not as low as in Profile 1, and did not have family risk for RD. Furthermore, Profile 5 had average cognitive skills, but below-average reading skills. Each individual's profile membership was saved.

Cross-tabulation and a chi-square test were conducted between the profile membership variable and the RF difficulty (Table 11) and RC difficulty (Table 12) variables using the testing data (the same 30% as in MLP). The proportions of RF difficulty differed between the latent profiles [ $\chi^2(7) = 23.51$ ,  $p < .01$ ]. Similarly, the proportions of RC difficulty differed between the latent profiles [ $\chi^2(7) = 36.53$ ,  $p < .001$ ]. The adjusted standardized residuals indicated that both RF difficulties (adj. residual 4.06) and RC difficulties (adj. residual 5.04) were more common in Profile 1 than they would have been by chance.

### Comparison of the three models

Next, we compared the three models' ability to predict RF or RC difficulties. The prediction scores and latent class information from the three models were used as independent variables and the dichotomous RF/RC variables as dependent in the logistic regression model. The chi-square test was used to reveal which model provided a prediction of reading difficulty that had additional information in relation to other models when trying to predict RD. The comparisons were performed using the 30% of data used for the testing model in the MLP.

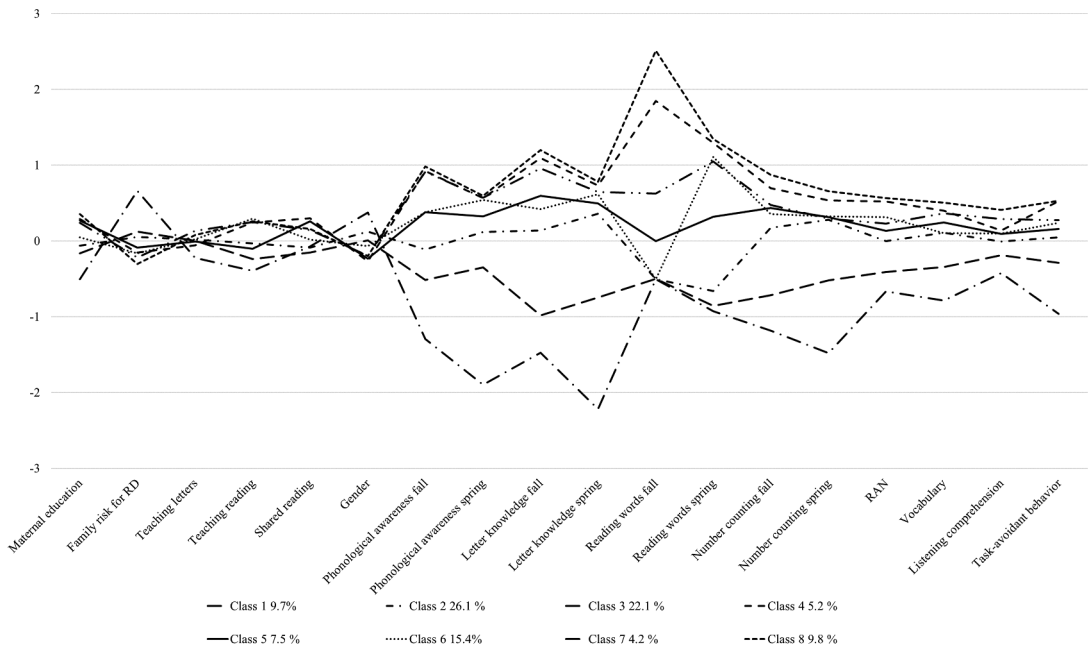
For RF, the chi-square test indicated that the MLP increased the prediction over BReg [ $\chi^2(1) =$

**Table 9.** Fit Indices for latent profile analyses.

Number of profiles	BIC	AIC	Entropy	p-value of LMR	p-value of VLMR
1	89,261.55	89,062.15			
2	82,544.30	82,239.66	.98	.00	.00
3	78,670.12	78,260.23	.95	.00	.00
4	76,007.68	75,492.55	.95	.00	.00
5	75,007.87	74,387.50	.94	.00	.00
6	74,198.21	73,472.60	.94	.00	.00
7	73,619.29	72,788.44	.97	.24	.23
8	72,746.11	71,810.02	.95	.00	.00
9	72,112.23	71,070.89	.95	.80	.80

BIC = Bayesian Information Criterion; LMR = Lo–Mendell–Rubin Adjusted Likelihood Ratio Test; VLMR = Vuong–Lo–Mendell–Rubin Likelihood Ratio Test

Lower values of BIC represent better model fit. The LMR and the VLMR compare the estimated model with the model having one fewer profile than the estimated model. A p-value of less than .05 shows that the estimated model is better and that the model with one fewer profile should be rejected. Entropy ranges from 0 to 1, and higher values show higher classification utility. In addition, the clarity of the latent profiles was examined by the average posterior probabilities for the most likely latent profile membership, which shows how distinct the profiles are.



**Figure 5.** Latent profiles of cognitive and home environment factors assessed in kindergarten.

21.46,  $p < .001$ ], while the BReg did not increase the prediction over MLP [ $\chi^2(1) = .58, p = .45$ ]. The MLP increased the prediction over LPA [ $\chi^2(1) = 17.70, p < .001$ ], while the LPA did not increase the prediction over MLP [ $\chi^2(7) = 6.13, p = .52$ ]. The LPA did not increase the prediction over the BReg [ $\chi^2(7) = 12.88, p = .08$ ]. Also, the BReg did not increase the prediction over LPA [ $\chi^2(1) = 3.57, p = .06$ ]. Furthermore, because the results indicated that the MLP was providing the most information, we examined whether the MLP could provide more information than the other two methods combined. The results indicated that the MLP increased the prediction over the LPA and the BReg combined [ $\chi^2(1) = 14.28, p < .001$ ].

For RC, the chi-square test indicated that the MLP increased the prediction over BReg [ $\chi^2(1) = 22.21, p < .001$ ], while the BReg did not increase the prediction over MLP [ $\chi^2(1) = .47, p = .49$ ]. The MLP increased the prediction over LPA [ $\chi^2(1) = 23.47, p < .001$ ], while the LPA did not increase the prediction over MLP [ $\chi^2(7) = 7.79, p = .35$ ]. The LPA increased the prediction over the BReg [ $\chi^2(7) = 16.58, p < .05$ ]. Also, the BReg increased the prediction over LPA [ $\chi^2(1) = 10.53, p < .01$ ]. Finally, the MLP increased the prediction over the LPA and the BReg combined [ $\chi^2(1) = 14.05, p < .001$ ].

In addition to the chi-square tests, the ROC curves for the BReg and MLP for RF (Figure 1) and RC (Figure 2) suggest that for both skills, MLP had greater power to predict RF and RC difficulties. For RF, the area under the curve for the MLP was .79, while for the BReg, it was .72. For RC, the area under the curve for the MLP was .83, while for the BReg, it was .75.

## Discussion

The complexity of risk and protective factor patterns in RD challenge the early identification of at-risk children. It is critical that we find methods that can handle the complexity of multiple interacting predictors to ensure that children are provided with timely support. The present study aimed to compare three models – linear (BReg), mixture (LPA), and neural networks (MLP) – based on their accuracy in predicting adolescent RD with multiple kindergarten-age variables. Furthermore, we examined whether the three models identified the same or different set of kindergarten-age predictors

**Table 10.** Means and standard errors for each profile for the kindergarten and home environment measures.

Measure	Profile 1 (N = 183)		Profile 2 (N = 489)		Profile 3 (N = 414)		Profile 4 (N = 97)		Profile 5 (N = 140)		Profile 6 (N = 293)		Profile 7 (N = 79)		Profile 8 (N = 185)		Comparisons between the profiles	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.		
Maternal education	-.52	.11	-.06	.06	-.17	.06	.30	.12	.28	.09	.05	.07	.25	.11	.36	.08	F(7, 1550) = 12.74***	1 < 2,3,4,5,6,7,8
Family risk for RD	.68	.16	.06	.06	.13	.08	-.16	.12	-.09	.11	-.17	.07	-.23	.13	-.31	.08	$\chi^2(7) = 61.21***$	More with family risk in profile 1 than would have been expected (adj. standardized 6.1)
Teaching letters	-.22	.09	.02	.06	.00	.07	-.06	.13	-.01	.08	.02	.06	.13	.14	.08	.08	F(7, 1560) = 1.38	
Teaching reading	-.41	.08	-.04	.05	-.25	.07	.25	.13	-.11	.09	.30	.07	.26	.12	.27	.08	F(7, 1561) = 12.61***	1 < 2,4,6,7,8
Shared reading	-.08	.10	-.09	.06	-.16	.06	.30	.10	.26	.09	.02	.06	.14	.11	.16	.08	F(7, 1551) = 4.87***	
Gender	.38	.08	.13	.05	.01	.06	-.27	.10	-.24	.09	-.07	.06	-.20	.11	-.20	.07	$\chi^2(7) = 60.44***$	More boys in profile 1 than would have been expected (adj. standardized 5.3)
Phonological awareness fall	-1.72	.16	-.15	.06	-.68	.08	1.22	.05	.50	.08	.50	.06	1.21	.05	1.30	.03	F(7, 1859) = 203.80***	1 < 2,3,4,5,6,7,8
Phonological awareness spring	-2.65	.41	.17	.06	-.49	.10	.79	.05	.45	.08	.75	.06	.79	.05	.83	.05	F(7, 1828) = 246.86***	1 < 2,3,4,5,6,7,8
Letter knowledge fall	-2.82	.09	.27	.10	-1.88	.10	2.08	.07	1.14	.11	.80	.09	1.83	.10	2.29	.06	F(7, 1859) = 739.69***	1 < 2,3,4,5,6,7,8
Letter knowledge spring	-4.94	.47	.80	.11	-1.67	.19	1.63	.12	1.11	.14	1.37	.11	1.45	.13	1.74	.12	F(7, 1828) = 1097.96***	1 < 2,3,4,5,6,7,8
Reading words fall	-7.11	.37	-7.11	.37	-7.10	.37	26.11	1.52	-.02	.00	-.71	.37	8.88	.41	35.48	1.83	F(7, 1859) = 52,972.60***	1 < 2,4,5,6,7,8
Reading words spring	-2.56	.08	-1.81	.08	-2.36	.09	3.56	.13	.88	.20	3.05	.13	2.88	.23	3.70	.15	F(7, 1815) = 1743.73***	1 < 2,3,4,5,6,7,8
Number counting fall	-1.53	.09	.22	.08	-.93	.09	.90	.10	.56	.10	.46	.07	.61	.12	1.13	.06	F(7, 1858) = 176.19***	1 < 2,3,4,5,6,7,8

(Continued)

Table 10. (Continued).

Measure	Profile 1 (N = 183)		Profile 2 (N = 489)		Profile 3 (N = 414)		Profile 4 (N = 97)		Profile 5 (N = 140)		Profile 6 (N = 293)		Profile 7 (N = 79)		Profile 8 (N = 185)		Comparisons between the profiles	
	Estimate	S.E	Estimate	S.E	Estimate	S.E	Estimate	S.E	Estimate	S.E	Estimate	S.E	Estimate	S.E	Estimate	S.E		
Number counting spring	-1.86	.18	.35	.06	-.65	.10	.67	.06	.39	.08	.41	.06	.37	.11	.82	.05	F(7, 1828) = 149.70***	1 < 2,3,4,5,6,7,8
RAN	-.72	.12	.00	.05	-.44	.06	.56	.08	.14	.07	.34	.05	.25	.14	.61	.07	F(7, 1827) = 43.20***	1 < 2,4,5,6,7,8
Vocabulary	-.85	.12	.12	.05	-.37	.06	.43	.10	.26	.08	.11	.06	.39	.12	.54	.07	F(7, 1831) = 41.69***	1 < 2,3,4,5,6,7,8
Listening comprehension	-.44	.10	-.01	.05	-.19	.06	.14	.11	.10	.08	.10	.06	.30	.14	.42	.07	F(7, 1824) = 13.16***	1 < 2,4,5,6,7,8
Task-avoidant behavior	-1.05	.11	.05	.05	-.31	.07	.58	.09	.17	.09	.26	.06	.30	.11	.58	.07	F(7, 1797) = 48.54***	1 < 2,3,4,5,6,7,8

For the variables family risk and gender a cross tabulation and a chi-square was used. For the rest of the variables a one-way ANOVA was conducted. Post-hoc comparisons were conducted using either Bonferroni or Dunnett T3, depending on equality of the variances. The LPA was conducted only with the kindergarten-age assessments and without the imputation of the data (N = 1,880). \*p < .05, \*\*p < .01, \*\*\*p < .001.

**Table 11.** Crosstable with the latent profiles and the reading fluency difficulties.

		Reading fluency		Total
		noRD	RD	
<i>Profile 1</i>	Count	16	7	23
	% within profile	69.57	30.43	100.00
	% within reading fluency	5.05	25.00	6.67
	% of Total	4.64	2.03	6.67
	Adjusted Residual	-4.06	4.06	
<i>Profile 2</i>	Count	80	10	90
	% within profile	88.89	11.11	100.00
	% within reading fluency	25.24	35.71	26.09
	% of Total	23.19	2.90	26.09
	Adjusted Residual	-1.21	1.21	
<i>Profile 3</i>	Count	57	6	63
	% within profile	90.48	9.52	100.00
	% within reading fluency	17.98	21.43	18.26
	% of Total	16.52	1.74	18.26
	Adjusted Residual	-.45	.45	
<i>Profile 4</i>	Count	23	0	23
	% within profile	100.00	.00	100.00
	% within reading fluency	7.26	.00	6.67
	% of Total	6.67	.00	6.67
	Adjusted Residual	1.48	-1.48	
<i>Profile 5</i>	Count	30	2	32
	% within profile	93.75	6.25	100.00
	% within reading fluency	9.46	7.14	9.28
	% of Total	8.70	.58	9.28
	Adjusted Residual	.41	-.41	
<i>Profile 6</i>	Count	63	2	65
	% within profile	96.92	3.08	100.00
	% within reading fluency	19.87	7.14	18.84
	% of Total	18.26	.58	18.84
	Adjusted Residual	1.65	-1.65	
<i>Profile 7</i>	Count	11	0	11
	% within profile	100.00	.00	100.00
	% within reading fluency	3.47	.00	3.19
	% of Total	3.19	.00	3.19
	Adjusted Residual	1.00	-1.00	
<i>Profile 8</i>	Count	37	1	38
	% within profile	97.37	2.63	100.00
	% within reading fluency	11.67	3.57	11.01
	% of Total	10.72	.29	11.01
	Adjusted Residual	1.31	-1.31	
<i>Total</i>	Count	317	28	345
	% within profile	91.88	8.12	100.00
	% within reading fluency	100.00	100.00	100.00
	% of Total	91.88	8.12	100.00
	Adjusted Residual			

The cross-tabulation between the LPA profile grouping and the dichotomized RF variable was conducted using the testing sample (the same 30% of the data used for the testing round in the MLP) (N = 345).

**Table 12.** Crosstable with the latent profiles and the reading comprehension difficulties.

		Reading comprehension		
		noRD	RD	Total
<i>Profile 1</i>	Count	11	10	21
	% within profile	52.38	47.62	100
	% within reading comprehension	4.14	26.32	6.91
	% of Total	3.62	3.29	6.91
	Adjusted Residual	-5.04	5.04	
<i>Profile 2</i>	Count	84	10	94
	% within profile	89.36	10.64	100
	% within reading comprehension	31.58	26.32	30.92
	% of Total	27.63	3.29	30.92
	Adjusted Residual	.66	-.66	
<i>Profile 3</i>	Count	49	12	61
	% within profile	80.33	19.67	100
	% within reading comprehension	18.42	31.58	20.07
	% of Total	16.12	3.95	20.07
	Adjusted Residual	-1.89	1.89	
<i>Profile 4</i>	Count	16	0	16
	% within profile	100	.00	100
	% within reading comprehension	6.02	.00	5.26
	% of Total	5.26	.00	5.26
	Adjusted Residual	1.55	-1.55	
<i>Profile 5</i>	Count	21	3	24
	% within profile	87.50	12.50	100
	% within reading comprehension	7.89	7.89	7.89
	% of Total	6.91	.99	7.89
	Adjusted Residual	.00	.00	
<i>Profile 6</i>	Count	46	3	49
	% within profile	93.88	6.12	100
	% within reading comprehension	17.29	7.89	16.12
	% of Total	15.13	.99	16.12
	Adjusted Residual	1.47	-1.47	
<i>Profile 7</i>	Count	10	0	10
	% within profile	100	.00	100
	% within reading comprehension	3.76	.00	3.29
	% of Total	3.29	.00	3.29
	Adjusted Residual	1.22	-1.22	
<i>Profile 8</i>	Count	29	0	29
	% within profile	100	.00	100
	% within reading comprehension	10.90	.00	9.54
	% of Total	9.54	.00	9.54
	Adjusted Residual	2.14	-2.14	
<i>Total</i>	Count	266	38	304
	% within profile	87.50	12.50	100
	% within reading comprehension	100	100	100
	% of Total	87.50	12.50	100
	Adjusted Residual			

The cross-tabulation between the LPA profile grouping and the dichotomized RC variable was conducted using the testing sample (the same 30% of the data used for the testing round in the MLP) (N = 304).

of Grade 9 RF and RC difficulties. The linear model was used as a basis because of its regular use in reading research. The mixture model is another often utilized method which was included because, unlike the linear model, it can identify the patterns or profiles of individuals among multiple variables that may signal heightened risk for RD. Neural networks models, although widely used in other fields, have been rarely used in reading research so far. The results indicated that MLP was more informative in predicting RF and RC difficulties than the other two models, a finding in line with the notion that neural networks are a powerful tool for predictive modeling.

### ***Accuracy in predicting RD***

All the models significantly predicted Grade 9 RF and RC scores with the kindergarten-age variables. LPA identified a group that had considerably higher risk for RD and both BReg and MLP predicted significant amounts of variance in the Grade 9 RD. Utilizing the same data employed for the testing model in MLP (30% of the data), the area under the curve for the MLP was .79 for RF and .83 for RC, while for the BReg, it was .72 for RF and .75 for RC. A value between .70 and .80 is considered acceptable, while a value between .80 and .90 is considered excellent (Mandrekar, 2010). Even though the results from the MLP and BReg were both considered acceptable (for RC, the MLP produced excellent results), it seems that the MLP classification was somewhat more accurate. Comparing the MLP, LPA, and BReg showed that neither LPA nor BReg provided any additional information for predicting either RF or RC difficulties over the MLP. In contrast, the MLP provided additional information over the other two methods, both for RF and RC. In fact, the MLP was found to be even more informative in predicting RF and RC difficulties than the combination of information from the other two models. Although, to the best of our knowledge, there are no previous studies comparing the use of these models for the prediction of RD, the findings were in line with our initial hypotheses expecting MLP to show high accuracy based on previous studies' findings in other fields that have used neural networks for prediction (e.g., Alzheimer's: Lu et al., 2018; ADHD: Kuang & He, 2014; autism: Heinsfeld et al., 2018; biomedicine: Mamoshina et al., 2016; Parkinson's: Choi et al., 2017; psychiatry: Vieira et al., 2017).

Linear models generally have been the dominant statistical models because their functions can be solved quickly, usually in one step, and their solutions are unique and represent a global optimum (i.e., the best possible solution among all solutions). Furthermore, the model fit works with relatively small samples, and the testing can be well-understood (Durstewitz et al., 2019). Compared with linear models, neural networks are more time-consuming and complicated, and it is often even impossible to identify the global optimum (Goodfellow et al., 2016). They also require large amounts of data for the training of the parameters (Durstewitz et al., 2019) and are more difficult to interpret (Urban & Gates, 2021). However, regardless of these difficulties, they can identify complex patterns in data and make very accurate predictions compared with linear models (Durstewitz et al., 2019). Furthermore, neural networks continue to work when the number of variables is larger than the sample size and in cases in which many weak factors correlated with each other influence the phenomenon (Urban & Gates, 2021). Another great benefit of the neural networks model is that the trained neural networks model can be generalized and used to perform the same task with new inputs (Cichy & Kaiser, 2019). In RD prediction models, as in ours, there are typically many kindergarten-age predictors which are correlated with each other. These correlations cause collinearity problems in the linear models but not in the MLP model. This may have contributed to the better performance of the MLP over the BReg. Another possible reason of why the MLP performed better than the BReg in the identification of RD is that the MLP can identify linear, non-linear, and interactional effects (for example, several combinations of the independent variables building cumulative risk) or a combination of these which the BReg cannot identify.

### **Important kindergarten factors for predicting adolescent RD**

Among the three models, some consensus was reached on which factors are the most important for predicting adolescent RF or RC difficulties. Unfortunately, while the MLP model provides information on which are the most important factors for identifying RD, it does not provide information on whether the effects of these factors are statistically significant. Overall, three factors emerged as the most important for predicting RF difficulties from the MLP: RAN, gender, and letter knowledge (spring). The same three factors, with the addition of number counting (spring), emerged as the most significant from the BReg. For RC, RAN, letter knowledge (spring), and vocabulary emerged as the most important factors from the MLP. From the BReg, RAN, gender, vocabulary, reading words (spring), and maternal education level were the most significant. The LPA indicated that both RF and RC difficulties were more common in Profile 1 than in the other profiles. In Profile 1, children had low scores in all kindergarten cognitive skills, and they also had family risk for RD. For the LPA, the comparison with the profiles found that those belonging to Profile 1 were male, had family risk for RD and had significantly lower scores in all cognitive skills than those belonging to the other profiles.

Both the MLP and BReg suggested that RAN was an important factor in predicting RF and RC difficulties. This is in line with previous studies that have reported RAN's significance as a predictor of RF (Georgiou et al., 2016; Landerl et al., 2019), particularly in more transparent orthographies such as Finnish (Eklund et al., 2018; Kairaluoma et al., 2013). Considering the association between word reading fluency skills and RC (see also, Florit & Cain, 2011), and that RAN is among the best predictors of reading fluency (e.g., Eklund et al., 2018; Georgiou et al., 2016; Landerl et al., 2019), it is not surprising that RAN also emerged as an important predictor of RC. Even though decoding's predictive power drops over time (e.g., Florit & Cain, 2011), previous studies have reported that RF remains a significant factor in predicting adolescent RC (García & Cain, 2014; Manu et al., 2021; Stanley et al., 2018; Torppa et al., 2018). Thus, it is likely that RAN's effect on RC found in this study is mediated via RF (e.g., Eklund et al., 2018). In addition to RAN, letter knowledge emerged as an important factor in predicting RF difficulties by both the MLP and BReg, which is in accordance with several previous studies (e.g., Clayton et al., 2020; Manu et al., 2021).

Interestingly, BReg and LPA suggested that number counting may also be an important factor in predicting RF difficulties, which is noteworthy considering that number-counting skills usually are not included in predicting RF difficulties. However, studies that have included number counting in the prediction of RF skills have suggested that it could be a significant additional predictor (Bernabini et al., 2021; Koponen et al., 2016). Finally, vocabulary was found to be an important factor in predicting RC difficulties by both the MLP and BReg. This is also in accordance with previous studies suggesting that RC is heavily reliant on oral language comprehension skills, such as vocabulary (e.g., Petscher et al., 2018; Psyridou et al., 2018). Weaknesses in oral language comprehension can manifest years before a child learns to read, suggesting a possible causal link from oral language skills to later RC difficulties (e.g., Eklund et al., 2018; Petscher et al., 2018; Psyridou et al., 2018).

Unlike previous studies, task-avoidant behavior (Georgiou et al., 2010; Hirvonen et al., 2010), shared parent-child book reading, and the teaching literacy skills at home (e.g., Khanolainen et al., 2020; Mol & Bus, 2011; Silinskas et al., 2020a) were not identified as important factors. Similarly, unlike previous studies (e.g., Snowling & Melby-Lervåg, 2016; Torppa et al., 2011), neither the MLP nor BReg found family risk to be a significant predictor of RF difficulties in our data. One possible reason, at least for family risk, may be that it was related to poor cognitive skills and the inclusion of cognitive skills in the predictive models hide its effect on RD. As the LPA indicates, those belonging to Profile 1 and had the lowest scores in cognitive skills in kindergarten also had family risk for RD. Another possible reason could be the measure used to assess family risk in this study: a child was considered as having family risk if either the mother or father reported some or clear RD. This assessment method is not as sensitive as a more formal assessment, considering that in another Finnish sample, parental RD has been found to predict RF over and above children's cognitive skills (Torppa et al., 2011). Family risk may have emerged as a predictor among those in Profile 1 from the LPA because the LPA focuses only on similarities and differences in respect to the kindergarten-age factors and does not include reading skills. Finally, both the



MLP and BReg identified gender as an important factor for RF and RC. Many previous studies have reported that females outperform males in reading and that more males are identified with RD than females (e.g., Psyridou et al., 2021; Quinn & Wagner, 2015).

This study has certain limitations that need to be addressed. First, we only used one measure to assess RC. A more extensive assessment could have yielded a more reliable assessment of RC and, subsequently, better predictions. Second, the assessment of listening comprehension was not optimal and the reliability estimate was quite low. Such low reliability could be one reason why listening comprehension did not emerge as an important factor for predicting RC difficulties. Thus, the results regarding the strongest predictors of RC should be interpreted with caution, as it is likely that they are underestimating listening comprehension's predictive value. Third, our sample had missing values. We balanced the imputed cases between the training and testing models by running the analysis multiple times, but not having missing values would have been optimal. The handling of missing cases did not follow the same approach across the three models. For the neural networks, we used single imputation (an expectation-maximization algorithm in SPSS) while for the linear model and the mixture model we used data without imputing the missing values. Imputation was conducted for the neural networks model because the MLP could not be performed with missing values. The linear and the mixture models on the other hand do not have such kind of limitation and can handle missingness, hence no need to use imputed data. For the linear and mixture models we used a full information maximum likelihood estimator (MLR estimator in Mplus). With this method we reach accurate standard errors while imputed data produces biases in standard errors. If we were to use imputed data for the linear and mixture models, we would have obtained equal parameter estimates to the MLR method used now but downward biased standard errors. This is because single imputation decreases the variability by adding the most plausible value to the data. Moreover, we think that it is important to use the way which is considered the best for handling missing data for each individual method as this is also the manner in which they are used in empirical papers that do not aim to compare methods against each other. Fourthly, the family risk variables were based on self-reports with a single question. This assessment may not have provided an accurate evaluation of parental reading skills and, thus, underestimating the predictive power of parental reading skills. In addition, RF or RC difficulty was defined as scoring in the lowest 10% of the RF or RC distribution, respectively. The selection of the cutoff matters as they are always somewhat arbitrary, and the outcomes of the studies might also be affected by the chosen cutoff (Psyridou et al., 2020). However, although the use of cutoffs is likely to lead to uncertainties in research findings because of measurement error (Branum-Martin et al., 2013; Francis et al., 2005; Psyridou et al., 2020), it is also a practical tool for the identification of children with RD. Internationally, different research groups use different cutoffs such as the lowest 10%, lowest 25%, 1SD below the mean, 1.5SD below the mean (e.g., Catts et al., 2012: 1SD, Esmaeeli et al., 2019: lowest 20%; Etmanskie et al., 2016: 25th percentile, Snowling et al., 2021, p. 10th percentile). The large sample of the present study allowed the selection of a rather strict cutoff for the identification of RD. Finally, in this study we compared three models on their accuracy in predicting RD and we identified that the MLP was more accurate than both the LPA and BReg models. Future research should examine how other models (e.g., random forest, LASSO regression) compare with neural networks models.

## Conclusions

Overall, the results of this study suggest that the neural networks model (MLP) was the most accurate method compared with the linear (BReg) and mixture (LPA) models for the early prediction of adolescent RF and RC difficulties. MLP was found to be superior even when compared with the combination of LPA and BReg. Thus, the results indicate that the use of deep neural networks for early RD identification purposes in the case of data with multiple interacting predictors looks promising. Possible reasons for the better performance of the neural networks could be that the neural networks models do not run into collinearity problems compared to the linear models. This allows the inclusion of a much broader set of predictors. In addition, the neural networks models can identify non-linear effects which the linear models cannot identify. This could also lead to the identification of more complex patterns in the data and as such

better performance of the model. For example, the neural network approach could handle many more variables than the ones we have used here, e.g., more information about home, preschool, early school, or non-cognitive psychological factors, which holds a promise for even better prediction of RD. After careful development of a trained model, this method maybe could be used ultimately in schools to aid in early identification of children at risk of developing RD.

With respect to the further use of neural networks for predicting RD, some practical and ethical considerations must be made. First, when training models with many parameters on small sample sizes, it can be difficult to identify a solution that can be generalized to the larger population (Whelan & Garavan, 2014). Training the model in different data sets may be necessary to make it more generalizable (Durstewitz et al., 2019). As [Appendices B and C](#) show, the correlation between the training and testing models for RF and RC changes depending on the part of the data used. Second, machine learning entails using a trained model to make predictions for new inputs; thus, training the model is critical because later predictions will be only as good as the data used for training. Biased data, intentionally or unintentionally, can lead to biased results (Ouchchy et al., 2020; Stahl, 2021). As researchers, we must be aware of the models and data used for training. Third, while this type of modeling opens the door to predicting RD early, indicating potential for early support, this early prediction also increases the risk of students encountering discrimination, labeling, and being left behind due to an algorithms result. Although neural networks seemed to provide a more accurate prediction, it was still far from being 100% accurate; therefore, false positive and false negative classifications still occurred. However, accuracy could be improved by adding more data. In this study, we predicted RF and RC difficulties at one time point (Grade 9) using kindergarten-age factors. Assessing reading skills at more than one time point, particularly during the early phases of reading development, likely would improve the accuracy in predicting adolescent RD. Furthermore, the inclusion of a broader and more reliable battery of measures might help improve predictions of RD. However, even if prediction accuracy were to rise sharply, in practice, the neural networks model should serve merely as one tool among many when early screening and support decisions are made concerning individuals. We also must remember that early identification is useful and ethically fully justified only if it is followed by adequate support.

In conclusion, neural networks seem to be a valid method for identifying those at risk for RD early. MLP was more accurate than both the LPA and BReg models. In the past, linear and mixture models have been used most often for this purpose, but machine learning models, such as the neural networks model described in this study, seem to be more powerful, exceeding both the linear and mixture models as well as their combination.


## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the Academy of Finland [Grant numbers #263891, #268586, #276239, #284439, #292466, #313768, #339418].

## ORCID

Maria Psyridou  <http://orcid.org/0000-0003-1425-2279>  
Asko Tolvanen  <http://orcid.org/0000-0001-6430-8897>  
Priyanka Patel  <http://orcid.org/0000-0002-1181-1066>  
Daria Khanolainen  <http://orcid.org/0000-0002-1571-2938>  
Marja-Kristiina Lerkkanen  <http://orcid.org/0000-0002-5709-5800>  
Anna-Maija Poikkeus  <http://orcid.org/0000-0001-7913-8691>  
Minna Torppa  <http://orcid.org/0000-0003-3834-9892>

## References

- Aro, M. (2017). Learning to read Finnish. In L. T. W. Verhoeven & C. A. Perfetti (Eds.), *Reading acquisition across languages and writing systems: An international handbook* (1st ed., pp. 416–436). Cambridge University Press.
- Aro, T., Eklund, K., Eloranta, A.-K., Närhi, V., Korhonen, E., & Ahonen, T. (2019). Associations between childhood learning disabilities and adult-age mental health problems, lack of education, and unemployment. *Journal of Learning Disabilities*, 52(1), 71–83. <https://doi.org/10.1177/0022219418775118>
- Asparouhov, T., & Muthen, B. (2008). Multilevel mixture models. In G. R. Hancock & K. M. Samuelson (Eds.), *Advances in latent variable mixture models* (pp. 27–51). Information Age Publishing.
- Bernabini, L., Bonifacci, P., & de Jong, P. F. (2021). The relationship of reading abilities with the underlying cognitive skills of math: A dimensional approach. *Frontiers in Psychology*, 12, 577488. <https://doi.org/10.3389/fpsyg.2021.577488>
- Branum-Martin, L., Fletcher, J. M., & Stuebing, K. K. (2013). Classification and identification of reading and math disabilities: The special case of comorbidity. *Journal of Learning Disabilities*, 46(6), 490–499. <https://doi.org/10.1177/0022219412468767>
- Caravolas, M., Lervåg, A., Mikulajová, M., Defior, S., Seidlová-Málková, G., & Hulme, C. (2019). A cross-linguistic, longitudinal study of the foundations of decoding and reading comprehension ability. *Scientific Studies of Reading*, 23(5), 386–402. <https://doi.org/10.1080/10888438.2019.1580284>
- Choi, H., Ha, S., Im, H. J., Paek, S. H., & Lee, D. S. (2017). Refining diagnosis of Parkinson’s disease with deep learning-based interpretation of dopamine transporter imaging. *NeuroImage: Clinical*, 16, 586–594. <https://doi.org/10.1016/j.nicl.2017.09.010>
- Cichy, R. M., & Kaiser, D. (2019). Deep neural networks as scientific models. *Trends in Cognitive Sciences*, 23(4), 305–317. <https://doi.org/10.1016/j.tics.2019.01.009>
- Clayton, F. J., West, G., Sears, C., Hulme, C., & Lervåg, A. (2020). A longitudinal study of early reading development: Letter-sound knowledge, phoneme awareness and RAN, but not letter-sound integration, predict variations in reading development. *Scientific Studies of Reading*, 24(2), 91–107. <https://doi.org/10.1080/10888438.2019.1622546>
- Denckla, M. B., & Rudel, R. G. (1976). Rapid “automatized” naming (R.A.N): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, 14(4), 471–479. [https://doi.org/10.1016/0028-3932\(76\)90075-0](https://doi.org/10.1016/0028-3932(76)90075-0)
- Dunn, L. M., & Dunn, L. M. (1981). *Peabody picture vocabulary test—Revised*. American Guidance Service.
- Durstewitz, D., Koppe, G., & Meyer-Lindenberg, A. (2019). Deep neural networks in psychiatry. *Molecular Psychiatry*, 24(11), 1583–1598. <https://doi.org/10.1038/s41380-019-0365-9>
- Eklund, K., Torppa, M., Sulkunen, S., Niemi, P., & Ahonen, T. (2018). Early cognitive predictors of PISA reading in children with and without family risk for dyslexia. *Learning and Individual Differences*, 64, 94–103. <https://doi.org/10.1016/j.lindif.2018.04.012>
- Erbeli, F., Hart, S. A., Wagner, R. K., & Taylor, J. (2018). Examining the etiology of reading disability as conceptualized by the hybrid model. *Scientific Studies of Reading*, 22(2), 167–180. <https://doi.org/10.1080/10888438.2017.1407321>
- Esmaeeli, Z., Kyle, F. E., & Lundström, K. (2019). Contribution of family risk, emergent literacy and environmental protective factors in children’s reading difficulties at the end of second-grade. *Reading and Writing*, 32(9), 2375–2399. <https://doi.org/10.1007/s11145-019-09948-5>
- Etmanskie, J. M., Partanen, M., & Siegel, L. S. (2016). A longitudinal examination of the persistence of late emerging reading disabilities. *Journal of Learning Disabilities*, 49(1), 21–35. <https://doi.org/10.1177/0022219414522706>
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Florit, E., & Cain, K. (2011). The simple view of reading: Is it valid for different types of alphabetic orthographies? *Educational Psychology Review*, 23(4), 553–576. <https://doi.org/10.1007/s10648-011-9175-6>
- Francis, D. J., Fletcher, J. M., Stuebing, K. K., Lyon, G. R., Shaywitz, B. A., & Shaywitz, S. E. (2005). Psychometric approaches to the identification of LD: IQ and achievement scores are not sufficient. *Journal of Learning Disabilities*, 38(2), 98–108. <https://doi.org/10.1177/00222194050380020101>
- García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*, 84(1), 74–111. <https://doi.org/10.3102/0034654313499616>
- Georgiou, G. K., Manolitsis, G., Nurmi, J. E., & Parrila, R. (2010). Does task-focused versus task-avoidance behavior matter for literacy development in an orthographically consistent language? *Contemporary Educational Psychology*, 35(1), 1–10. <https://doi.org/10.1016/j.cedpsych.2009.07.001>
- Georgiou, G. K., Parrila, R., & Papadopoulos, T. C. (2016). The anatomy of the RAN-reading relationship. *Reading and Writing*, 29(9), 1793–1815. <https://doi.org/10.1007/s11145-016-9653-9>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. <http://www.deeplearningbook.org>
- Graves, A., Mohamed, A. R., & Hinton, G. (2013, May). Speech recognition with deep recurrent neural networks. In 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 6645–6649). <https://doi.org/10.1109/ICASSP.2013.6638947>

- Hamilton, L. G., Hayiou-Thomas, M. E., Hulme, C., & Snowling, M. J. (2016). The home literacy environment as a predictor of the early literacy development of children at family-risk of dyslexia. *Scientific Studies of Reading, 20*(5), 401–419. <https://doi.org/10.1080/10888438.2016.1213266>
- Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., & Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage: Clinical, 17*, 16–23. <https://doi.org/10.1016/j.nicl.2017.08.017>
- Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive Sciences, 11*(10), 428–434. <https://doi.org/10.1016/j.tics.2007.09.004>
- Hirvonen, R., Georgiou, G. K., Lerkkanen, M. K., Aunola, K., & Nurmi, J. E. (2010). Task-focused behaviour and literacy development: A reciprocal relationship. *Journal of Research in Reading, 33*(3), 302–319. <https://doi.org/10.1111/j.1467-9817.2009.01415.x>
- Kairaluoma, L., Torppa, M., Westerholm, J., Ahonen, T., & Aro, M. (2013). The nature and factors related to reading difficulties among adolescents in a transparent orthography. *Scientific Studies of Reading, 5*(5), 315–332. <https://doi.org/10.1080/10888438.2012.701257>
- Khanolainen, D., Psyridou, M., Silinskas, G., Lerkkanen, M. K., Niemi, P., Poikkeus, A. M., & Torppa, M. (2020). Longitudinal effects of the home learning environment and parental difficulties on reading and math development across grades 1–9. *Frontiers in Psychology, 11*, 2735. <https://doi.org/10.3389/fpsyg.2020.577981>
- Koponen, T., Aunola, K., Ahonen, T., & Nurmi, J. E. (2007). Cognitive predictors of single-digit and procedural calculation skills and their covariation with reading skill. *Journal of Experimental Child Psychology, 97*(3), 220–241. <https://doi.org/10.1016/j.jecp.2007.03.001>
- Koponen, T., Salmi, P., & Nurmi, J.-E. (2016). Counting and rapid naming predict the fluency of arithmetic and reading skills. *Contemporary Educational Psychology, 44*, 83–94. <https://doi.org/10.1016/j.cedpsych.2016.02.004>
- Kuang, D., & He, L. (2014, November). Classification on ADHD with deep learning. In *2014 International Conference On Cloud Computing And Big Data* (pp. 27–32). IEEE.
- Landerl, K., Freudenthaler, H. H., & Georgiou, G. K. (2019). Phonological awareness and rapid automatized naming as longitudinal predictors of reading in five alphabetic orthographies with varying degrees of consistency. *Scientific Studies of Reading, 23*(3), 220–234. <https://doi.org/10.1080/10888438.2018.1510936>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lerkkanen, M.-K., Poikkeus, A.-M., & Ketonen, R. (2006). *ARMI—Lukuja kirjoitustaidon arviointimateriaali 1: Luokalle*. WSOY.
- Lerkkanen, M.-K., Eklund, K., Löytynoja, H., Aro, M., & Poikkeus, A. M. (2018). *YKÄ Lukuja kirjoitustaidon arviointimenetelmä yläkouluun*. [YKÄ—Reading test for lower secondary school]. Niilo Mäki Instituutti.
- Lindeman, J. (2000). *ALLU—Ala-asteen lukutesti* [ALLU reading test for primary school]. University of Turku, Centre for Research on Learning.
- Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198–1202. <https://doi.org/10.1080/01621459.1988.10478722>
- Lu, D., Popuri, K., Ding, G. W., Balachandar, R., & Beg, M. F. (2018). Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer’s disease using structural MR and FDG-PET images. *Scientific Reports, 8*(1), 1–13. <https://doi.org/10.1038/s41598-018-22871-z>
- Mamoshina, P., Vieira, A., Putin, E., & Zhavoronkov, A. (2016). Applications of deep learning in biomedicine. *Molecular Pharmaceutics, 13*(5), 1445–1454. <https://doi.org/10.1021/acs.molpharmaceut.5b00982>
- Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology, 5*(9), 1315–1316. <https://doi.org/10.1097/JTO.0b013e3181ec173d>
- Manu, M., Torppa, M., Eklund, K., Poikkeus, A. M., Lerkkanen, M. K., & Niemi, P. (2021). Kindergarten pre-reading skills predict grade 9 reading comprehension (PISA reading) but fail to explain gender difference. *Reading and Writing, 34*(3), 753–771. <https://doi.org/10.1007/s11145-020-10090-w>
- McArthur, G., Badcock, N., Castles, A., & Robidoux, S. (2021). Tracking the relations between children’s reading and emotional health across time: Evidence from four large longitudinal studies. *Reading Research Quarterly, 57*(2), 555–585. <https://doi.org/10.1002/rrq.426>
- 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>
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., . . . Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature, 518*(7540), 529–533. <https://doi.org/10.1038/nature14236>
- 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. <https://doi.org/10.1037/a0021890>
- Nation, K., Cocksey, J., Taylor, J. S., & Bishop, D. V. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry, 51*(9), 1031–1039. <https://doi.org/10.1111/j.1469-7610.2010.02254.x>
- Nevala, J., & Lyytinen, H. (2000). *Sanaketjutesti*. Niilo Mäki Instituutti & Jyväskylän yliopiston lapsitutkimuskeskus. Differentiate word test

- OECD. (2010). *PISA 2009 results: Learning trends: Changes in student performance since 2000*. <https://doi.org/10.1787/9789264091580-en>
- Ouchchy, L., Coin, A., & Dubljević, V. (2020). AI in the headlines: The portrayal of the ethical issues of artificial intelligence in the media. *AI & SOCIETY*, 35(4), 927–936. <https://doi.org/10.1007/s00146-020-00965-5>
- 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>
- Petscher, Y., Justice, L. M., & Hogan, T. (2018). Modeling the early language trajectory of language development when the measures change and its relation to poor reading comprehension. *Child Development*, 89(6), 2136–2156. <https://doi.org/10.1111/cdev.12880>
- Psyridou, M., Eklund, K., Poikkeus, A. M., & Torppa, M. (2018). Reading outcomes of children with delayed early vocabulary: A follow-up from age 2–16. *Research in Developmental Disabilities*, 78, 114–124. <https://doi.org/10.1016/j.ridd.2018.05.004>
- Psyridou, M., Tolvanen, A., Lerkkanen, M. K., Poikkeus, A. M., & Torppa, M. (2020). Longitudinal stability of reading difficulties: Examining the effects of measurement error, cut-offs, and buffer zones in identification. *Frontiers in Psychology*, 10, 2841. <https://doi.org/10.3389/fpsyg.2019.02841>
- Psyridou, M., Tolvanen, A., de Jong, P. F., Lerkkanen, M. K., Poikkeus, A. M., & Torppa, M. (2021). Developmental profiles of reading fluency and reading comprehension from grades 1 to 9 and their early identification. *Developmental Psychology*, 57(11), 1840. <https://doi.org/10.1002/rrq.426>
- Puglisi, M. L., Hulme, C., Hamilton, L. G., & Snowling, M. J. (2017). The home literacy environment is a correlate, but perhaps not a cause, of variations in children's language and literacy development. *Scientific Studies of Reading*, 21(6), 498–511. <https://doi.org/10.1080/10888438.2017.1346660>
- Quinn, J. M., & Wagner, R. K. (2015). Gender differences in reading impairment and in the identification of impaired readers: Results from a large-scale study of at-risk readers. *Journal of Learning Disabilities*, 48(4), 433–445. <https://doi.org/10.1177/0022219413508323>
- Ricketts, J., Lervåg, A., Dawson, N., Taylor, L. A., & Hulme, C. (2020). Reading and oral vocabulary development in early adolescence. *Scientific Studies of Reading*, 24(5), 380–396. <https://doi.org/10.1080/10888438.2019.1689244>
- Sanfilippo, J., Ness, M., Petscher, Y., Rappaport, L., Zuckerman, B., & Gaab, N. (2020). Reintroducing dyslexia: Early identification and implications for pediatric practice. *Pediatrics*, 146(1). <https://doi.org/10.1542/peds.2019-3046>
- Schatschneider, C., Wagner, R. K., Hart, S. A., & Tighe, E. L. (2016). Using simulations to investigate the longitudinal stability of alternative schemes for classifying and identifying children with reading disabilities. *Scientific Studies of Reading*, 20(1), 34–48. <https://doi.org/10.1080/10888438.2015.1107072>
- Schiefele, U., Schaffner, E., Möller, J., & Wigfield, A. (2012). Dimensions of reading motivation and their relation to reading behavior and competence. *Reading Research Quarterly*, 47(4), 427–463. <https://doi.org/10.1002/RRQ.030>
- Sénéchal, M., Lefevre, J. A., Thomas, E. M., & Daley, K. E. (1998). Differential effects of home literacy experiences on the development of oral and written language. *Reading Research Quarterly*, 33(1), 96–116. <https://doi.org/10.1598/RRQ.33.1.5>
- Silinskas, G., Sénéchal, M., Torppa, M., & Lerkkanen, M.-K. (2020a). Home literacy activities and children's reading skills, independent reading, and interest in literacy activities from kindergarten to Grade 2. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.01508>
- Silinskas, G., Torppa, M., Lerkkanen, M. K., & Nurmi, J. E. (2020b). The home literacy model in a highly transparent orthography. *School Effectiveness and School Improvement*, 31(1), 80–101. <https://doi.org/10.1080/09243453.2019.1642213>
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., . . . Hassabis, D. (2016). Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587), 484–489. <https://doi.org/10.1038/nature16961>
- Snowling, M. J., & Melby-Lervåg, M. (2016). Oral language deficits in familial dyslexia: A meta-analysis and review. *Psychological Bulletin*, 142(5), 498–545. <https://doi.org/10.1037/bul0000037>
- Snowling, M. J., Nash, H. M., Gooch, D. C., Hayiou-Thomas, M. E., & Hulme, C., & Wellcome Language and Reading Project Team. (2019). Developmental outcomes for children at high risk of dyslexia and children with developmental language disorder. *Child Development*, 90(5), e548–e564. <https://doi.org/10.1111/cdev.13216>
- Snowling, M. J., Moll, K., & Hulme, C. (2021). Language difficulties are a shared risk factor for both reading disorder and mathematics disorder. *Journal of Experimental Child Psychology*, 202, 105009. <https://doi.org/10.1016/j.jecp.2020.105009>
- Spencer, M., Wagner, R. K., Schatschneider, C., Quinn, J. M., Lopez, D., & Petscher, Y. (2014). Incorporating RTI in a hybrid model of reading disability. *Learning Disability Quarterly*, 37(3), 161–171. <https://doi.org/10.1177/0731948714530967>
- Stahl, B. C. (2021). Ethical Issues of AI. *Artificial Intelligence for a Better Future*, 35. [https://doi.org/10.1007/978-3-030-69978-9\\_4](https://doi.org/10.1007/978-3-030-69978-9_4)
- Stanley, C. T., Petscher, Y., & Catts, H. (2018). A longitudinal investigation of direct and indirect links between reading skills in kindergarten and reading comprehension in tenth grade. *Reading and Writing*, 31(1), 133–153. <https://doi.org/10.1007/s11145-017-9777-6>

- Statistics Finland. (2007). *Statistical databases*. Retrieved March 1, 2019, from [www.stat.fi/tup/tilastotietokannat/index\\_en.html](http://www.stat.fi/tup/tilastotietokannat/index_en.html)
- Torppa, M., Eklund, K., van Bergen, E., & Lyytinen, H. (2011). Parental literacy predicts children's literacy: A longitudinal family risk study. *Dyslexia*, 17(4), 339–355. <https://doi.org/10.1002/dys.437>
- Torppa, M., Eklund, K., Sulkunen, S., Niemi, P., & Ahonen, T. (2018). Why do boys and girls perform differently on PISA reading in Finland? The effects of reading fluency, achievement behaviour, leisure reading and homework activity. *Journal of Research in Reading*, 41(1), 122–139. <https://doi.org/10.1111/1467-9817.12103>
- Urban, C. J., & Gates, K. M. (2021, April 1). Deep learning: A primer for psychologists. *Psychological Methods*, 26(6), 743–773. <https://dx.doi.org/10.1037/met0000374>.
- 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>
- Vauras, M., Mäki, H., Dufva, M., & Hämäläinen, S. (1995). Diagnostiset testit 2. Kuullun ja luetun ymmärtäminen. In *Listening and reading comprehension*. Diagnostic tests 2. Oppimistutkimuksen keskus, Turun yliopisto.
- Vieira, S., Pinaya, W. H., & Mechelli, A. (2017). Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. *Neuroscience and Biobehavioral Reviews*, 74, 58–75. <https://doi.org/10.1016/j.neubiorev.2017.01.002>
- Whelan, R., & Garavan, H. (2014). When optimism hurts: Inflated predictions in psychiatric neuroimaging. *Biological Psychiatry*, 75(9), 746–748. <https://doi.org/10.1016/j.biopsych.2013.05.014>
- Zhang, X., Nurmi, J.-E., Kiuru, N., Lerkkanen, M.-K., & Aunola, K. (2011). A teacher-report measure of children's task-avoidant behavior: A validation study of the behavioral strategy rating scale. *Learning and Individual Differences*, 21(6), 690–698. <https://doi.org/10.1016/j.lindif.2011.09.007>

## Appendix A

Correlations among the kindergarten-age factors and the reading fluency and reading comprehension skills in Grade 9

	1	2	3	4	5	6	7	8	9	10	11
1. Maternal education	1	-.20**	-.09**	.00	.22**	.04	.21**	.17**	.24**	.18**	.20**
2. Family risk for RD	-.20**	1	-.04	-.08*	-.07*	.04	-.13**	-.16**	-.19**	-.18**	-.14**
3. Teaching letters	-.08**	-.04	1	.59**	.14**	-.09**	-.02	-.01	.06*	.05*	.02
4. Teaching reading	.00	-.07*	.60**	1	.09**	-.11**	.08**	.14**	.17**	.18**	.10**
5. Shared reading	.22**	-.07*	.14**	.11**	1	-.03	.11**	.06*	.15**	.10**	.12**
6. Gender	.04	.04	-.09**	-.11**	-.03	1	-.11**	-.16**	-.12**	-.12**	-.13**
7. Phonological awareness fall	.20**	-.12**	-.01	.08**	.11**	-.10**	1	.56**	.61**	.54**	.56**
8. Phonological awareness spring	.17**	-.17**	.01	.13**	.04	-.17**	.55**	1	.55**	.59**	.37**
9. Letter knowledge fall	.23**	-.19**	.06*	.19**	.15**	-.13**	.58**	.52**	1	.81**	.63**
10. Letter knowledge spring	.18**	-.19**	.07**	.20**	.08**	-.15**	.52**	.62**	.79**	1	.50**
11. Reading words fall	.17**	-.14**	.02	.12**	.10**	-.11**	.46**	.28**	.55**	.37**	1
12. Reading words spring	.18**	-.19**	.02	.21**	.11**	-.15**	.57**	.50**	.67**	.60**	.62**
13. Number counting fall	.18**	-.14**	-.03	.10**	.00	.13**	.42**	.38**	.60**	.55**	.38**
14. Number counting spring	.15**	-.16**	.01	.08**	-.01	.09**	.37**	.40**	.50**	.55**	.29**
15. RAN	.13**	-.14**	.03	.07**	.08**	-.05*	.26**	.28**	.34**	.32**	.25**
16. Vocabulary	.20**	-.08**	.00	.04	.19**	.00	.32**	.29**	.34**	.29**	.22**
17. Listening comprehension	.13**	-.10**	.02	.05	.14**	-.07**	.18**	.16**	.18**	.16**	.16**
18. Task avoidant behavior	.16**	-.13**	-.01	.07**	.01	-.24**	.31**	.31**	.34**	.36**	.24**
19. Word list reading	.17**	-.06	.01	.09**	.06	-.24**	.22**	.21**	.32**	.32**	.26**
20. Differentiate words	.15**	-.13**	.02	.12**	.05	-.34**	.24**	.25**	.34**	.35**	.30**
21. Sentence reading fluency	.15**	-.14**	.05	.12**	.12**	-.24**	.20**	.19**	.27**	.29**	.25**
22. Reading fluency mean score	.18**	-.12**	.03	.13**	.09**	-.32**	.26**	.25**	.36**	.37**	.31**
23. Reading comprehension PISA	.23**	-.11**	.00	.09*	.21**	-.22**	.29**	.23**	.32**	.30**	.28**

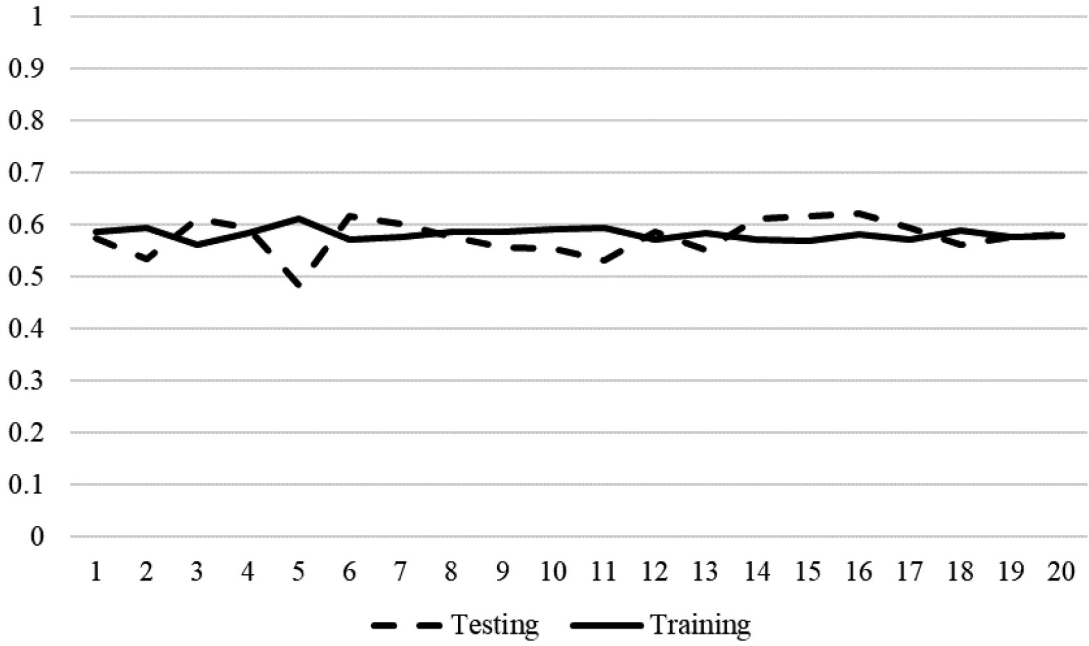
(continue)

	12	13	14	15	16	17	18	19	20	21	22	23
1. Maternal education	.17**	.18**	.15**	.14**	.20**	.14**	.17**	.18**	.15**	.16**	.18**	.22**
2. Family risk for RD	-.19**	-.15**	-.14**	-.12**	-.07*	-.08**	-.14**	-.06	-.14**	-.14**	-.12**	-.11**
3. Teaching letters	.03	-.03	.00	.02	-.02	.02	.00	.01	.01	.05	.03	-.01
4. Teaching reading	.20**	.09**	.05*	.07**	.03	.04	.07*	.08**	.12**	.12**	.12**	.07*
5. Shared reading	.10**	.00	.00	.09**	.19**	.15**	.02	.05	.04	.12**	.08**	.21**
6. Gender	-.17**	.13**	.09**	-.08**	.01	-.07**	-.24**	-.24**	-.34**	-.24**	-.32**	-.22**
7. Phonological awareness fall	.62**	.43**	.35**	.29**	.30**	.19**	.30**	.23**	.24**	.22**	.26**	.29**
8. Phonological awareness spring	.63**	.39**	.35**	.31**	.26**	.14**	.30**	.23**	.27**	.21**	.27**	.24**
9. Letter knowledge fall	.71**	.59**	.48**	.37**	.32**	.19**	.33**	.32**	.35**	.28**	.36**	.33**
10. Letter knowledge spring	.72**	.56**	.51**	.37**	.27**	.15**	.31**	.32**	.36**	.31**	.37**	.30**
11. Reading words fall	.62**	.41**	.31**	.28**	.24**	.17**	.26**	.27**	.31**	.26**	.32**	.33**
12. Reading words spring	1	.50**	.41**	.38**	.25**	.17**	.35**	.31**	.39**	.31**	.39**	.37**
13. Number counting fall	.48**	1	.66**	.30**	.26**	.15**	.30**	.27**	.28**	.23**	.29**	.25**
14. Number counting spring	.39**	.69**	1	.30**	.24**	.11**	.28**	.22**	.26**	.21**	.26**	.20**
15. RAN	.33**	.30**	.31**	1	.17**	.14**	.26**	.30**	.26**	.30**	.32**	.24**
16. Vocabulary	.24**	.28**	.27**	.19**	1	.30**	.20**	.15**	.07*	.18**	.14**	.28**
17. Listening comprehension	.17**	.15**	.14**	.11**	.29**	1	.16**	.11**	.11**	.13**	.13**	.23**
18. Task avoidant behavior	.33**	.30**	.32**	.25**	.22**	.17**	1	.23**	.27**	.22**	.27**	.24**
19. Word list reading	.31**	.27**	.24**	.27**	.17**	.11**	.21**	1	.64**	.69**	.88**	.43**
20. Differentiate words	.39**	.28**	.26**	.22**	.07*	.12**	.26**	.66**	1	.59**	.85**	.42**
21. Sentence reading fluency	.31**	.23**	.21**	.27**	.19**	.13**	.21**	.67**	.57**	1	.87**	.50**
22. Reading fluency mean score	.39**	.30**	.27**	.29**	.16**	.13**	.27**	.89**	.86**	.86**	1	.51**
23. Reading comprehension PISA	.36**	.25**	.21**	.23**	.30**	.22**	.24**	.43**	.43**	.48**	.52**	1

Below the diagonal there is Pearson correlation coefficient and above the diagonal Spearman's  $\rho$ . \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$

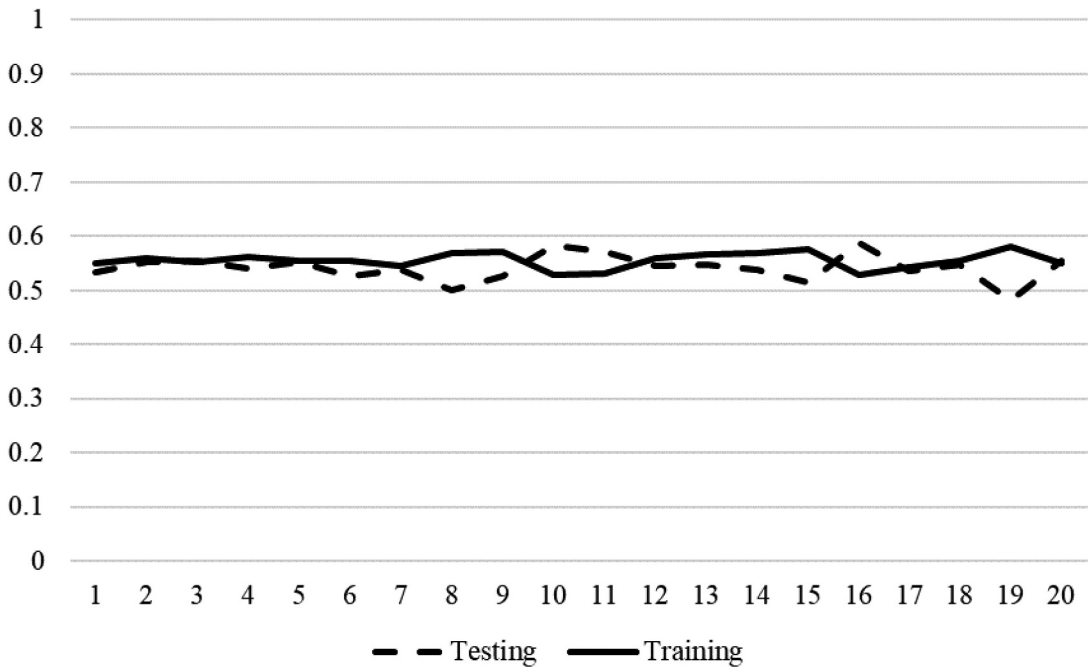
### Appendix B

Correlations between the training and the testing model for the 20 seeds used in the MLP model for reading fluency



### Appendix C

Correlations between the training and the testing model for the 20 seeds used in the MLP model for reading comprehension





## Appendix D

Estimates and the p-values at each step of the multiple regression analysis with backward deletion for the prediction of reading fluency difficulties

	Step 1		Step 2		Step 3		Step 4		Step 5		Step 6		Step 7		Step 8	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Maternal education	-.16	.29	-.15	.31	-.16	.28	-.16	.28	-.17	.25	-.17	.24	-.17	.25	-.15	.28
Family risk for RD	-.12	.45	-.11	.49	-.11	.50	-.09	.56	-.11	.49	-.11	.48	-.10	<b>.53</b>	*	*
Teaching letters	.07	.67	.07	.66	.07	.66	.07	<b>.66</b>	*	*	*	*	*	*	*	*
Teaching reading	-.12	.44	-.13	.41	-.12	.45	-.13	.43	-.10	.49	-.09	<b>.50</b>	*	*	*	*
Shared reading	-.05	.71	-.04	<b>.78</b>	*	*	*	*	*	*	*	*	*	*	*	*
Gender	.86	.00	.86	.00	.86	.00	.86	.00	.85	.00	.85	.00	.85	.00	.84	.00
Phonological awareness fall	.21	.19	.20	.21	.20	.22	.21	.17	.21	.17	.22	.16	.23	.14	.22	.15
Phonological awareness spring	.05	.72	.05	.74	.04	<b>.74</b>	*	*	*	*	*	*	*	*	*	*
Letter knowledge fall	.19	.40	.19	.41	.19	.41	.18	.42	.18	.42	.19	.41	.17	.45	.17	<b>.46</b>
Letter knowledge spring	-.42	.03	-.42	.03	-.42	.03	-.40	.03	-.40	.04	-.41	.03	-.41	.03	-.40	.04
Reading words fall	-.28	.23	-.29	.22	-.27	.24	-.28	.23	-.27	.24	-.28	.23	-.26	.25	-.26	.25
Reading words spring	-.17	.40	-.16	.42	-.17	.40	-.16	.42	-.17	.39	-.16	.41	-.18	.35	-.18	.36
Number counting fall	-.14	.34	-.14	.34	-.14	.35	-.14	.35	-.15	.31	-.14	.33	-.14	.33	-.14	.35
Number counting spring	-.23	.15	-.23	.15	-.23	.15	-.22	.15	-.22	.15	-.22	.17	-.22	.17	-.22	.17
RAN	-.32	.00	-.33	.00	-.33	.00	-.32	.00	-.33	.00	-.33	.00	-.32	.00	-.32	.00
Vocabulary	-.02	<b>.88</b>	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Listening comprehension	.06	.65	.06	.67	.05	.68	.06	.65	.05	<b>.66</b>	*	*	*	*	*	*
Task-avoidant behavior	-.19	.15	-.19	.16	-.18	.18	-.18	.18	-.18	.18	-.18	.17	-.18	.19	-.17	.20

(continue)

	Step 9		Step 10		Step 11		Step 12		Step 13		Step 14		Step 15	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Maternal education	-.14	.31	-.15	.28	-.15	<b>.29</b>	*	*	*	*	*	*	*	*
Family risk for RD	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Teaching letters	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Teaching reading	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Shared reading	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Gender	.84	.00	.83	.00	.83	.00	.81	.00	.80	.00	.81	.00	.84	.00
Phonological awareness fall	.23	.13	.23	.14	.20	.19	.19	<b>.22</b>	*	*	*	*	*	*
Phonological awareness spring	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Letter knowledge fall	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Letter knowledge spring	-.31	.03	-.33	.02	-.37	.00	-.38	.00	-.33	.01	-.38	.00	-.40	.00
Reading words fall	-.23	.31	-.23	.29	-.31	.13	-.32	.12	-.27	<b>.18</b>	*	*	*	*
Reading words spring	-.16	.40	-.17	<b>.37</b>	*	*	*	*	*	*	*	*	*	*
Number counting fall	-.11	<b>.44</b>	*	*	*	*	*	*	*	*	*	*	*	*
Number counting spring	-.22	.15	-.27	.05	-.27	.05	-.28	.04	-.26	.05	-.27	.05	-.31	.02
RAN	-.31	.00	-.31	.00	-.32	.00	-.33	.00	-.32	.00	-.34	.00	-.36	.00
Vocabulary	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Listening comprehension	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Task-avoidant behavior	-.18	.19	-.18	.18	-.18	.17	-.20	.13	-.19	.16	-.19	<b>.14</b>	*	*

The standardized estimates are presented for each factor at each step along with the p-value. The factor with the highest p-value got deleted in the next step. This factor is presented with bold. Asterisks represent factors that have been deleted in previous steps. The BReg was conducted with the full sample without the imputation excluding the testing sample (N = 1,535).

## Appendix E

Estimates and the p-values at each step of the multiple regression analysis with backward deletion for the prediction of reading comprehension difficulties

	Step 1		Step 2		Step 3		Step 4		Step 5		Step 6		Step 7	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Maternal education	-.38	.02	-.37	.02	-.37	.02	-.38	.01	-.37	.01	-.38	.01	-.37	.01
Family risk for RD	.11	.52	.11	.54	.10	.55	.11	.53	.09	<b>.60</b>	*	*	*	*
Teaching letters	.04	.83	.03	<b>.84</b>	*	*	*	*	*	*	*	*	*	*
Teaching reading	-.21	.22	-.21	.23	-.20	.22	-.19	.22	-.18	.27	-.18	<b>.26</b>	*	*
Shared reading	-.19	.19	-.20	.17	-.19	.18	-.19	.19	-.19	.17	-.20	.15	-.22	.12
Gender	.62	.00	.62	.00	.62	.00	.61	.00	.62	.00	.62	.00	.62	.00
Phonological awareness fall	-.09	.60	-.09	.60	-.09	.60	-.09	<b>.60</b>	*	*	*	*	*	*
Phonological awareness spring	.30	.06	.30	.06	.30	.06	.30	.06	.28	.07	.28	.07	.27	.08
Letter knowledge fall	.43	.11	.43	.11	.44	.10	.44	.10	.43	.11	.43	.10	.41	.11
Letter knowledge spring	-.48	.03	-.49	.03	-.49	.03	-.50	.02	-.50	.02	-.51	.02	-.53	.02
Reading words fall	-.33	.27	-.32	.27	-.33	.27	-.32	.28	-.33	.26	-.33	.26	-.31	<b>.29</b>
Reading words spring	-.35	.13	-.36	.12	-.36	.12	-.36	.12	-.39	.08	-.39	.08	-.41	.06
Number counting fall	-.15	.40	-.15	.39	-.16	.37	-.17	.26	-.18	.24	-.19	.21	-.19	.20
Number counting spring	-.03	.84	-.03	.82	-.03	<b>.83</b>	*	*	*	*	*	*	*	*
RAN	-.36	.01	-.36	.01	-.36	.01	-.37	.01	-.36	.01	-.36	.01	-.35	.01
Vocabulary	-.51	.00	-.51	.00	-.51	.00	-.52	.00	-.53	.00	-.52	.00	-.52	.00
Listening comprehension	-.21	.11	-.21	.10	-.21	.10	-.21	.09	-.21	.09	-.22	.07	-.21	.08
Task-avoidant behavior	-.01	<b>.93</b>	*	*	*	*	*	*	*	*	*	*	*	*

(continue)

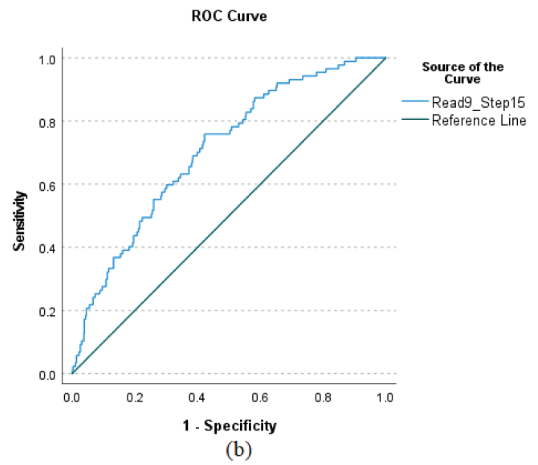
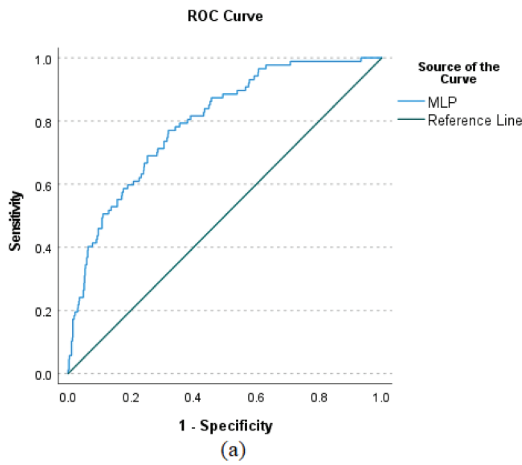


	Step 8		Step 9		Step 10		Step 11		Step 12		Step 13		Step 14	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Maternal education	-.38	.01	-.38	.01	-.38	.01	-.41	.00	-.41	.01	-.42	.00	-.43	.00
Family risk for RD	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Teaching letters	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Teaching reading	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Shared reading	-.22	.12	-.22	.12	-.23	<b>.10</b>	*	*	*	*	*	*	*	*
Gender	.63	.00	.59	.00	.59	.00	.59	.00	.56	.00	.58	.00	.59	.00
Phonological awareness fall	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Phonological awareness spring	.28	.07	.27	.08	.26	.09	.26	<b>.10</b>	*	*	*	*	*	*
Letter knowledge fall	.35	.17	.29	<b>.25</b>	*	*	*	*	*	*	*	*	*	*
Letter knowledge spring	-.49	.02	-.53	.01	-.35	.02	-.35	.02	-.25	<b>.08</b>	*	*	*	*
Reading words fall	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Reading words spring	-.52	.01	-.54	.01	-.46	.02	-.48	.02	-.42	.03	-.57	.00	-.60	.00
Number counting fall	-.19	<b>.20</b>	*	*	*	*	*	*	*	*	*	*	*	*
Number counting spring	*	*	*	*	*	*	*	*	*	*	*	*	*	*
RAN	-.35	.01	-.36	.01	-.34	.01	-.34	.01	-.32	.02	-.33	.01	-.32	.01
Vocabulary	-.52	.00	-.54	.00	-.51	.00	-.54	.00	-.52	.00	-.53	.00	-.58	.00
Listening comprehension	-.22	.08	-.23	.06	-.23	.07	-.25	.04	-.21	.07	-.23	<b>.05</b>	*	*
Task-avoidant behavior	*	*	*	*	*	*	*	*	*	*	*	*	*	*

The standardized estimates are presented for each factor at each step along with the p-value. The factor with the highest p-value got deleted in the next step. This factor is presented with bold. Asterisks represent factors that have been deleted in previous steps. The BReg was conducted with the full sample without the imputation excluding the testing sample (N = 1,576)

## Appendix Z

ROC Curves for reading fluency using MLP (figure a) and BReg (figure b) for the training models. The area under the curve for the MLP is .80 and for the BReg is .71. In the training sample there were 807 cases.



ROC Curves for reading comprehension using MLP (figure c) and BReg (figure d) for the training models. The area under the curve for the MLP is .84 and for the BReg is .72. In the training sample there were 711 cases.

