



Predicting Word Reading Deficits Using an App-Based Screening Tool at School Entry

Martin Schöfl^{1,2*}, Gabriele Steinmair¹, Daniel Holzinger^{2,3} and Christoph Weber^{1,2}

¹ Department of Educational Sciences, University of Education Upper Austria, Linz, Austria, ² Research Institute for Developmental Medicine, Johannes Kepler University Linz, Linz, Austria, ³ Institute of Neurology of Senses and Language, Hospital of St. John of God, Linz, Austria

OPEN ACCESS

Edited by:

Maria Pia Bucci,
Centre National de la Recherche
Scientifique (CNRS), France

Reviewed by:

Adrian Pasquarella,
University of Delaware, United States
Stephanie Ducrot,
Aix-Marseille Université, France

*Correspondence:

Martin Schöfl
Martin.Schoeffl@jku.at

Specialty section:

This article was submitted to
Children and Health,
a section of the journal
Frontiers in Pediatrics

Received: 27 January 2022

Accepted: 01 April 2022

Published: 11 May 2022

Citation:

Schöfl M, Steinmair G,
Holzinger D and Weber C (2022)
Predicting Word Reading Deficits
Using an App-Based Screening Tool
at School Entry.
Front. Pediatr. 10:863477.
doi: 10.3389/fped.2022.863477

Background: Reading is a crucial competence associated with academic development, mental health, and social adaptation. Reading difficulties are often detected at a late stage, with a possible negative impact on long-term reading development and secondary developmental disadvantages. The first manifestations of reading difficulties can be identified by word reading deficits in first and second grade, paving the way for specific interventions. For widespread implementation, instruments must be easy to use and motivating for children.

Objectives: Development and validation of an economical, well-accepted, and accurate screening tool composed of the domains of phonological information processing, language skills, and non-verbal intelligence in regular school settings.

Design: In 2020, the screening tool was used on a sample of 409 first graders between the second and fifth weeks of school in a one-to-one setting. Additionally, information on parental education and the use of German and/or other languages by the child was collected using a parental questionnaire. A follow-up involving the use of established standardized word reading tests was conducted at the end of the first school year.

Results: A five-variable screening tool consisting of the dimensions of phonological information processing (letter knowledge, rapid naming, and phonological awareness) and linguistic skills (receptive vocabulary and morphosyntax) showed statistical relevance (AUC = 0.78; sensitivity 0.80, specificity 0.74) for predicting word reading problems concerning reading speed (< 16th percentile) at the end of first grade, whereas gender, first language, and age of first exposure to the German language did not contribute to the prediction. The instrument was well accepted by the children and screeners and can be administered within an acceptable time frame.

Conclusion: Word reading deficits at the end of first grade can be predicted by the use of an app-based screening tool at school entry that includes phonological information processing and language skills. Further validation and assessment of empirical feasibility data are needed to support the screening instrument for German orthography.

Keywords: word reading, primary school, predictive power, language, app-based screening

INTRODUCTION

Word reading consists of several components: phonological analysis of the written word (1); orthographic processing in the sense of the “ability to form, store, and access orthographic representations” [(2), p. 4049], and lexical access to word meaning. In German orthography, which is characterized by low complexity and high transparency, reading difficulties manifest at an early stage as reduced word reading speed (3–10).

In an established reading model by Perfetti (11) and Perfetti and Hart (12), fluent word decoding and fast and effortless access to the orthographic lexicon predicted sentence and text reading fluency and thus represents an essential component of academic learning (13).

Reading weaknesses are associated with significant disadvantages throughout the school years and beyond, with impacts on school achievement (14, 15) and later employment (16). Nordström et al. (17) stressed the importance of schools investigating children’s early word decoding ability. Another large-scale longitudinal study (18) followed up individuals who had weak word decoding at the age of 7, finding that they had lower school achievement and income as adults compared to good and average readers. Stanovich summarized the logic of this finding as early as 1986 as the “Matthew Effect”: Good readers are intrinsically motivated to read and therefore read a lot, consequently, their reading skills continuously improve. Children starting school with poor reading skills often lack the motivation to read and consequently read less. Soon, a gap begins to open that is increasingly difficult to close.

Longitudinal studies have shown this very development in different orthographies [for German: (4, 5, 19, 20)].

An Australian research team conducted a comprehensive review of over 100 articles investigating the emotional consequences of slow reading in children over a period of 30 years, finding an increasingly negative impact on self-esteem and anxiety (21). German scientists replicated the findings of increased internalization of problems and resulting social withdrawal in children with reading or spelling deficits compared to children without learning disorders (22). Mammarella et al. (23) found that sustained academic failure and perceived low self-esteem increased the risk of anxiety and depression in children with reading problems. Earlier detection of risk factors connected with specific interventions could counteract this trend of reading deficits with consequences for education, employment, and wellbeing (24).

Screening tools [c.f. (25)] need to meet the following criteria: to be stable over time; to accurately predict reading achievement (criterion: “validity”); and to be objectively applicable, evaluable, and interpretable (criterion: “objectivity”). Additional criteria relate to their application in schools: screening must require little training of the test instructors; their administration should be time-efficient and administrable with limited staff resources (criterion: “test economy”). Furthermore, screening should be motivating and not overburdening for the children (criterion: “reasonableness”), and no child should be disadvantaged by

the way it is conducted or the language used (criterion: “fairness”). Finally, the results should be available to teachers quickly and unambiguously and should allow conclusions to be drawn for schools, such as assignments to support groups or the adaptation of teaching methods (criterion: “usefulness”).

The use of app-based screening technology by children around school-entry age appears promising in terms of both test economy and feasibility, as demonstrated by the assessment of vocabulary performance in the last year of kindergarten (26) in Austria. Internationally, acceptable clinical screening accuracy is reported only close to or at school entry and not in the prior years (27). This is because data on the highly relevant and directly literacy-related factors (e.g., letter knowledge) can only be collected close to the beginning of school entry, in addition to more general predictors, such as non-verbal intelligence or linguistic skills.

Child-Related Predictors

Linguistically based skills on the one hand and visual skills on the other hand have been found to predict word reading [for an overview, see (28)]. Predictors associated with visual processing such as visual memory span at kindergarten age (29) have been researched experimentally, but to our knowledge, there is still a lack of established test paradigms shown to be feasible within school-based screenings. Related to linguistically based predictors in alphabetic languages, letter knowledge, phonological awareness, and Rapid Automatized Naming-speed (RAN) have been demonstrated as robust predictors of word reading even across different orthographies and a number of reliable and practicable test paradigms have been developed (30). These factors, often summarized as phonological information processing (31, 32), are frequently supplemented by phonological working memory (33). Only recent studies have focused on the prerequisites for these factors, namely, linguistic skills. In a longitudinal study, Snowling et al. (34) demonstrated the influence of lower levels of linguistic competencies on the development of specific learning disorders. Moreover, linguistic deficits in already-diagnosed reading problems are observed retrospectively by parents more often than in average or good readers (35). Thus, asking parents about their children’s language performance and assessing it as an additional factor at school entry is anticipated to be a central component of a valid screening tool for written language skills. Non-verbal IQ as a general predictor of school success contributed little to direct variance explanation of word reading or writing performance in previous longitudinal studies. Rather, non-verbal IQ determined the level of profiles in profile analyses, such as the large-scale study by Ozernov-Palchik et al. (36). Non-verbal IQ as a general predictor of school success contributed little to direct variance explanation of word reading or writing performance in previous longitudinal studies. Rather, non-verbal IQ determined the level of profiles in profile analyses, such as the large-scale study by Ozernov-Palchik et al. (36). Latent profile analysis in kindergarteners showed specific effects and interactions of the known predictors RAN, phonological awareness, verbal working memory, and letter knowledge. The level of (non-verbal) IQ helped to identify

groups of children with average or overall (non-specific) slightly below-average performance. With regard to screening IQ as an additional criterion does not lead to better identification of children that require specific reading promotion and is therefore not included.

Environmental Factors

Environmental factors influencing children's reading competence have been highlighted in established reading socialization models [e.g., the multilevel model of family reading Hurrelmann et al. (37)]. They describe learners' reading experiences in different social contexts and their influence on the development of motivation, interests, and skills. Rosebrock and Nix's [(38), p. 16] reading literacy model includes three levels of reading competence: subjective, cognitive, and social (39). At the social level, the influence of the family as language and reading role models is emphasized as a moderator variable.

Lack of familiarity with the language spoken at school might be another factor that could affect word reading due to underspecified phonological representations or irregular letter-sound correspondence for L2 (40). Growing up with a primary language (L1) other than that used at school (L2) is usually related to having an immigrant background, including culture-specific home environments related to literacy. Although the majority of studies have found similar word reading skills in these children compared to their native peers (41–43), only a few reported better outcomes for native students (44, 45). For Dutch, there is a body of evidence in support of word decoding from kindergarten being highly comparable in Dutch as the first language (L1) and Dutch as the second language (L2) learners (46). Nevertheless, group differences at different stages of reading development have been documented, such as differences between L1 and L2 learners on rapid naming assessments in Grade 1, which disappeared in Grade 2 (47). For German (48), German L2 learners' reading fluency was mostly predicted by non-verbal intelligence, whereas for L1 learners, phonological awareness tasks in the last year of kindergarten best predicted reading fluency.

Scientific Aim

The aim was to construct and validate a time-efficient screening tool for word reading ability for use in community school settings around school entry. Child-related predictors concerned phonological information processing (phonological awareness, letter knowledge, and phonological working memory), language (vocabulary and grammar), and non-verbal intelligence. Children's gender, additional environmental predictors, first language, and exposure to the German language were analyzed as potential moderating variables.

MATERIALS AND METHODS

Participant Recruitment

The majority of the recruited children were from an Upper Austrian district with four big community-based schools. They were informed and invited to participate in the study project firstly via telephone and then by personal visits. All headmasters

agreed to participate. Additionally, four more schools asked for participation and joined the project. Finally, parents of 409 children (100%) gave their written permission for including their children in the study. The final study sample reflected a heterogeneous distribution of children, comparable with Austrian primary schools in terms of gender, the proportion of children with a non-German dominant language, and parental educational levels.

The individualized screening started in autumn 2020 within 2 weeks after school onset. Within 3 weeks, 85% of the sample had been assessed. In the two subsequent weeks, those children who had been ill or unable to attend during the first survey period were surveyed. A total of 27 children were not included in the analysis because they were listed as "Vorschüler" (preschoolers) in their first year of learning. A total of 86 children were sick or out of school when the reading test was conducted at the end of the first school year; these children did not differ significantly from the analyzed sample in terms of age, gender, and most importantly, the screening variables. At the end of first grade, word reading was assessed in the classroom setting. **Figure 1** shows the recruitment pathways and timeline.

Participant Characteristics

In school statistics for the 2019/2020 school year, of the 344,282 Austrian elementary school children, 48.2% were female (49). Nearly the same proportion is found in the present study, where 48.9% ($N = 407$) are female.

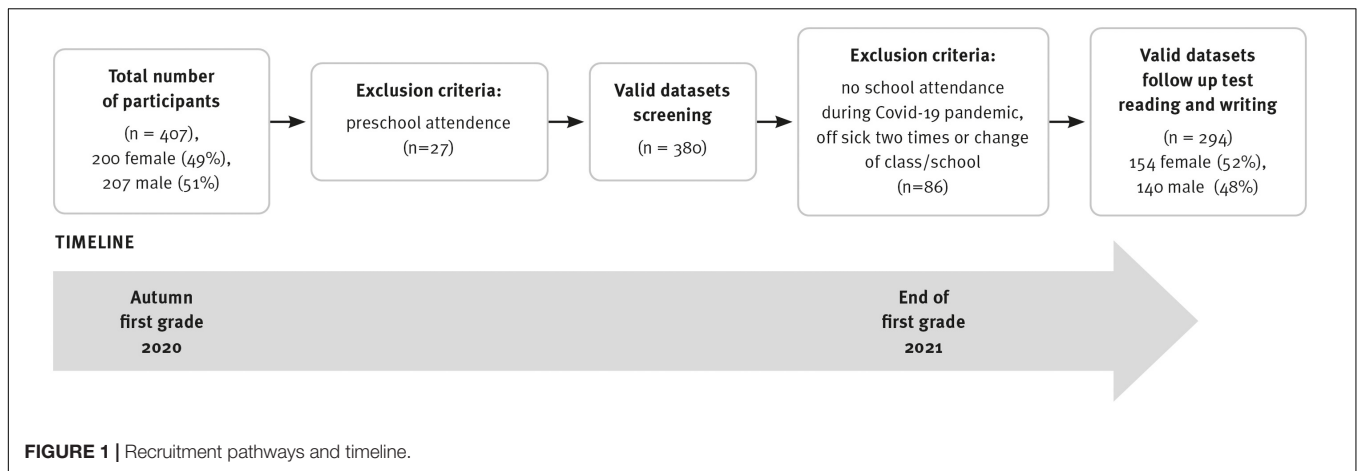
In the 2019/2020 school year, 106,498 out of a total of 344,282 children in Austrian elementary schools had German not as their first language, which corresponds to 26.8% (49). Hence, around seven out of ten children in Austrian elementary schools have German as their first language [(49), p. 12]. In all nine schools studied in the project, 74.2% of children speak German exclusively as a first language. The proportions from the research project thus correspond to the Austrian distribution.

Socioeconomic status is approximated by parents' highest educational attainment. The sample consisted of parents from all educational backgrounds: among the mothers, 4.9% had maximum educational attainment of an elementary school diploma, 14.4% had a high school diploma, and 29.8% had a university diploma. Educational levels for fathers were comparable (6.9% with maximum educational attainment of a primary school diploma, 15.4% with a high school diploma, and 21.5% with a university degree). Overall, the educational level of the Austrian population are comparable: 6% of parents have the highest educational attainment in elementary school, 22% in high school, and 27% have a university diploma. The given sample contains a variety of educational levels.

Measures and Procedure

This 1-year prospective study followed children from the beginning to the end of first grade. The research design included two steps:

- (1) Screening of phonological information processing, intelligence, and language in the first weeks of first grade, before the formal teaching of reading and spelling, had



begun. The classification of the first spoken language was done by means of a questionnaire by the parents. If it was indicated that the first language was only German or contact with German occurred from birth up to and including the age of 2, then children were classified as L1. Children whose contact with German occurred only after the age of 2 were classified as L2.

- (2) Standardized assessment of word reading at the end of first grade.

Screening Measures

The screening tool consisted of 13 subtests, which can be systematized into the three domains of phonological information processing, language, and non-verbal intelligence. A total of seven of the subtests are well-established standardized tests, whereas six of the screening tests have been newly designed; see **Table 1** for an overview of the tasks. All tasks were app-supported, although for some subtests the child had a paper version to look at or an audio presentation of the stimuli was played to them on a tablet, which was used to enter results.

Phonological Information Processing

Phonological awareness was assessed by three tasks, with one task intended to differentiate in the lowest performance range (rhyming), one in the middle (syllable count), and one close to written language acquisition (initial phoneme detection). Phonological awareness tasks were newly constructed despite the presence of existing tests in order to meet the quality criterion of the economy for the use of the instrument in the school setting. Existing test procedures in German-speaking countries are well constructed but are too time-consuming for universal use in schools [e.g., (50)]. Each of the tasks was introduced by three practice items including feedback, followed by ten test items. Tasks were constructed from high-frequency words from the childLex database (51) for the youngest age group (6–8 years). Syllable count was controlled for the target items, and distractor tasks consisted of phonologically similar structures to the target items.

For the rhyming task, the child selected the words that rhyme from a set of three words (picture and word presented). Ten examples (7 one-syllable and 3 two-syllable) were presented. For example: “What rhymes: house, mouse, man?” An explorative factor analysis (EFA) for binary items conducted with Mplus 8 (52) showed a dominant factor with an eigenvalue of 5.351 (53% explained variance). A second factor with an eigenvalue of 1.418 was not interpretable. Moreover, the one-factor solution yielded an acceptable fit [$\chi^2(35) = 0.9405$, root mean square error of approximation (RMSEA) = 0.065; comparative fit index (CFI) = 0.918]. Therefore, we choose the single-factor solution. Internal consistency was (Kuder–Richardson KR-20) low but acceptable at 0.610. Internal consistency was (Kuder–Richardson KR-20) low but acceptable at 0.61.

Syllable count was also assessed by one- to four-syllable words, presented by means of pictures and spoken language; visual cues (one clapping a hand to four clapping hands) were used to indicate the number of syllables. Notably, an EFA yielded an inadequate fit for a one-factor solution [$\chi^2(35) = 211.4$, RMSEA = 0.112, CFI = 0.850], but a good fit for a two-factor solution [$\chi^2(26) = 38.9$, RMSEA = 0.035, CFI = 0.989]. The analyses revealed a factor focusing on two or more syllable words (7 items; eigenvalue = 4.563, 46% explained variance) and a factor focusing on one-syllable words (3 items; eigenvalue = 1.876, 19% explained variance). Thus, we used two different syllable count scores in this paper. Internal consistency was good for the one-syllable factor (KR-20 = 0.764) and acceptable for the two or more-syllable factor (KR-20 = 0.651).

For initial phoneme detection, we presented a letter visually and as a speech sound simultaneously.

From a selection of three pictures, those with the same first phoneme had to be selected (“Which word begins with I like Ines: Hase, Igel, Spiegel?”). Although an EFA yielded three factors with eigenvalues greater than 1 (3.772, 1.418, 1.064), the one-factor solution showed an adequate fit [$\chi^2(35) = 74.3$, RMSEA = 0.053, CFI = 0.908]. Thus, we choose the single-factor solution. Internal consistency was adequate (KR-20 = 0.632).

For the assessment of Rapid Automated Naming (RAN), two conditions were chosen: objects and digits. The RAN

TABLE 1 | Subtests and domains of the screening measures.

Domain	Subtest	Type of subtest	Number of practice items	Number of test items	Presentation mode	Target selection mode	
Phonological information processing	Phonological awareness	Newly designed	3	10	Tablet	Children using tablet	
	Rhyme detection						
	Phonological awareness syllable count	Newly designed	3	10	Tablet	Children using tablet	
	Phonological awareness Initial phoneme detection	Newly designed	3	10	Tablet	Children using tablet	
	Rapid Automatized Naming, RAN (1) Objects	Newly designed	5	30	Paper	Instructor	
	Rapid Automatized Naming, RAN (2) Digits	Denckla and Rudel (53) following Landerl et al. (54)	5	30	Paper	Instructor	
	Letter knowledge	Newly designed	None	26	Paper	Instructor	
	Phonological working memory Word list memory	Newly designed	None	Adaptive	Tablet	Instructor	
	Phonological working memory	IDS-II, Grob and Hagmann-von Arx (55)	None	Adaptive	Instructor	Instructor	
	Letter-number-span forward						
	Phonological working memory	IDS-II, Grob and Hagmann-von Arx (55)	None	Adaptive	Instructor	Instructor	
	Letter-number-span backward						
	Linguistic skills	Receptive vocabulary	GraWo; Seifert et al. (58)	2	30	Tablet	Children using tablet
		Sentence repetition	Adapted from Hamann and Abend Ibrahim (57)	None	15	Tablet	Instructor
Intelligence	Complete matrices	PITVA (59)	None	Adaptive	Paper	Instructor	
	Picture series	PITVA (59)	None	Adaptive	Paper	Instructor	

object condition was designed through five high-frequency monosyllabic words (cow, hand, ice, tree, and mouse). First, the items of the RAN tasks were presented app-based, and the task was given to repeat these items. Once the investigator ensured that the instruction was understood and the items were known, the test session started. The items were presented on paper repeatedly in a different order over six lines. The investigator pressed a button on the tablet to time the test and noted any incorrect responses on the tablet by pressing a button. When the last item was reached, the time measurement was stopped manually again, and the time distance was calculated automatically.

Rapid Automatized Naming in the digit condition was based on the work of Denckla and Rudel (53), following (54), and was presented and rated analogously to the object condition with monosyllabic digits (2, 8, 1, 6, 3).

Letter knowledge: All letters of the alphabet were offered as capital letters in random order on paper. Each page contained three to four letters. Children were asked: "I know you haven't learned these letters yet at school. Maybe you still know one? Please name it!" Positive scores were given for letter names or sounds and ticked off on the tablet.

Phonological working memory was assessed by two subtests of a broad-range intelligence test battery [IDS-2; Intelligence and Development Scales for Children and Adolescents, (55)] testing memory of letter-number sequences forward and backward. The

child was asked to repeat a series of digits mixed up with letters (3-A, 5-M-2) in the same (forward condition) way or form back and forth (backward condition). The investigator clicked correct solutions on the tablet. The difficulty level of the tasks was determined by the length of the spans, and the termination occurred after three unsolved or incorrectly solved tasks. The longest possible range of letter and number sequences was of the target value. Reliability is described as fair; Cronbach's Alpha was 0.89 (end of first grade). Retest reliability was $r_{tt} = 0.93$ (first grade).

Wordlist memory was also used to analyze phonological working memory: a list of 10 words (5 single-syllable words and 5 two-syllable words) was presented *via* an audio file. The child was then asked to freely reproduce as many of them as possible. The investigator ticked off the words in the mentioned order (including repetitions and wrong words). The sum of all memorized items yielded the overall performance.

Language

Morphosyntactic skills were assessed by an adapted sentence-repetition task. The German version was constructed according to the LITMUS (Language Impairment Testing in Multilingual Children) principles (56) by Hamann et al. (57) following the COST Action IS0804. A block of 15 items representing morphosyntactic constructions with varying degrees of complexity was selected and scored according to whether or not the sentence was completed correctly. Correctness was

judged and noted in the app by the examiner when the sentence structure was reproduced completely correctly, regardless of articulatory deficits.

Internal consistency (KR-20) was high at 0.877.

In a digital form of the Graz Vocabulary Test [GraWo; (58)], receptive vocabulary was tested by 30 matching tasks. The child was required to choose from four pictures the one that matched the audio-presented word. Reliability data are given for the paper form of the GraWo: Cronbach's Alpha ranged from 0.89 (end of first grade) to 0.82 (end of second grade). Retest reliability was $r_{tt} = 0.93$ (first grade).

Non-verbal Intelligence

Two subtests of the PITVA (59) were used to assess non-verbal intelligence: Complete Matrices (Cronbach's Alpha 0.83 for 6 year old/Retest reliability $r_{tt} = 0.9$) and Picture Series (Cronbach's Alpha 0.86). The child was shown matrices and sequences of items and had to click on the correct condition from a selection directly on the tablet.

Word Reading

Word reading and writing tasks were administered in a classroom setting at the end of the first school year, exclusively by research staff.

The ELFE II word reading test (60) was used to assess decoding fluency at the word level in silent reading. For each picture, the appropriate written word from a selection of four had to be selected. The test duration was limited to 3 min. Representative norm scores are available from the end of the first school year to the beginning of the seventh grade; reliability data are presented as excellent (split-half $r = 0.98$, retest $r = 0.83$). A cut-off score of 13 represents $M - 1$ SD.

Procedure

Before the implementation of the screening tool, the principals of participating schools received information about the testing process. They were also given a letter to send to parents, including consent forms and questions about children's first language and language use as well as parents' educational background.

Teachers entered children's names into an online database, which converted the names to IDs for use on the tablet. The testers, all of them were student teachers, were enrolled in a student seminar in order to learn about the tasks and testing procedure, through which the teachers received student credit for the study (amounting to 4 h). The materials for testing were brought to the schools by a research coordinator. On the test mornings, it was agreed with the school administration that the children would be selected alphabetically by the test team (student teachers and core study project staff) from the classrooms. The assessment took place one at a time, with the child and instructor seated across a table from each other. After a brief welcome, the child was handed the tablet, in which the friendly dragon SCHWUPP was introduced right at the beginning. The app navigation was designed in a way that the child can use it independently, but if necessary, the test leader intervened in the navigation of the dragon from

one task to the next. All instructions essential for the child were recorded as audio files, opened automatically, and could be repeated if necessary. A yellow background on the app signaled to the test administrator that the child was making test selections independently (such as in the phonological tasks). A gray background meant that the test administrator had to take the tablet to read the instructions from the tablet and give the corresponding instructions. This was especially true for tasks with material (for example, the letter cards). The assessment including all subtests took an average of 38.4 min (SD = 9.3) per child.

METHODS OF ANALYSIS

First, we used receiver operator characteristic (ROC) analyses to evaluate the diagnostic accuracy of each subscale. Following Swets (61), AUCs ≥ 0.9 are regarded as excellent, AUCs ≥ 0.8 and < 0.9 as good, AUCs ≥ 0.7 and < 0.8 as fair, and tests with AUCs < 0.7 as poor. ROC analyses were conducted using the pROC package (62) in R.

Second, to construct a time-efficient screening to predict word reading difficulties at the end of Grade 1, we used a logistic regression model with adaptive variable selection to identify important subtests. In detail, we applied the least absolute shrinkage and selection operator [LASSO; e.g., (63, 64); for an application of LASSO for the selection of screening variables, see (65)] as implemented in the glmnet R Package (66), which adequately addresses the problem of overfitting that is pronounced in standard variable selection procedures (e.g., backward or forward selection) and models with many predictors (relative to the sample size). Overfitting occurs when sample regression estimates capture signal and noise and thus are larger than in the population, which in turn limits the generalizability of the regression results. LASSO addresses overfitting and consequently increases generalizability by applying a penalty term (λ) to the likelihood function that protects estimates from inflation. Just as in the backward or forward selection, null predictors are zeroed out (i.e., they are excluded from the prediction model). Notably, LASSO does not provide p -values (methods have been developed for linear models, but not for logistic models) and thus it is not possible to refer to the "significance" of predictors. Instead, the selected predictors are meaningful whether their effects are significantly different from zero or not (63, 65).

To evaluate the importance of the screening variables, we z -scored the predictors. Thus, reported estimates are in a standardized metric. Moreover, LASSO requires the selection of an appropriate penalty term. We used 10-fold cross-validation and selected the value for λ that resulted in the highest area under the curve (AUC). Since in some cases this may insufficiently address the problem of overfitting, we also report results for the second value of λ by applying the one standard error rule [i.e., selecting the largest value of λ at which the AUC is within one standard error of the largest AUC; see e.g., (64), p. 216]. Once a set of predictors had been selected, we used the regression coefficients of the LASSO models to estimate the probability of

scoring within the 10%-percentile of the word reading test at the end of Grade 1. This probability score (ranging from 0 to 1) was subsequently used as a screening measure.

Third, we compared the screening scores based on the results for λ at the maximum AUC with scores based on the one standard error rule by applying a bootstrapped test that compares the AUCs of paired ROC curves (62).

Fourth, we compared ROC curves between groups defined by the first language (German vs. non-German), German language exposure (≤ 2 years vs. > 2 years), and gender (girls vs. boys). Significantly differing ROC curves between groups indicate variations in diagnostic accuracy, which would consequently limit the generalizability of the screening tool (67). Besides using a bootstrapped test for unpaired ROC curves that compare the AUCs for two groups, we also applied the Venkatraman permutation test (68) that compares actual ROC curves (also implemented in the pROC package). If two ROC curves do not differ significantly between groups, screening scores would yield the same sensitivity and specificity in both groups, and thus, a single cut-off for both groups would be appropriate.

Finally, we determined optimal cutoff scores using the R-OptimalCutpoints package (69). Cut-offs were evaluated based on the following diagnostic accuracy statistics: sensitivity (Se), specificity (Sp), positive predictive values (PPV), negative predictive values (NPV), and diagnostic likelihood ratios for positive and negative screening results (DLR+ and DLR–, respectively). Se and Sp ≥ 0.9 indicate good diagnostic accuracy, Se and Sp ≥ 0.80 are regarded as fair, and values below 0.80 indicate an unacceptably high rate of misclassification (70). DLR+ and DLR– are diagnostic accuracy measures

that—unlike predictive values—do not depend on the prevalence of the disorder under investigation (71). DLR+ displays the multiplicative change in the pre-screening odds of scoring in the 10%-percentile of the reading test given a positive screening result (i.e., post-screening odds = DLR+ \times pre-screening odds). DLR– is the change in the pre-screening odds of scoring in the 10%-percentile given a negative screening result (post-screening odds = DLR– \times pre-screening odds). DLR+ values ≥ 10 and DLR– ≤ 0.1 indicate large changes in pre-screening odds, DLR+ ≤ 10 and > 5 , and DLR– > 0.1 and ≤ 0.2 indicate moderate changes, DLR+ ≤ 5 and > 2 , and DLR– > 0.2 and ≤ 0.5 indicate small changes. DLR+ < 2 and DLR– > 0.5 are rarely important (72).

RESULTS

Table 2 (Section A) shows the AUCs as well as the point-biserial correlations (r_{pb}) for the screening subtests. Notably, rhyme detection, syllable count, letter–number sequences forward, and the IQ subtests are not significantly associated with word reading problems at the end of Grade 1. For all other predictors, correlations are small and only the AUCs for RAN (digits and objects) and letter knowledge could be regarded as fair. Overall, the AUC for RAN objects is largest at 0.726 (DeLong 95%-CI [0.658, 0.795]), directly followed by letter knowledge (0.723, DeLong 95%-CI [0.645, 0.801]).

The results of the LASSO logistic regression models are reported in **Table 2** (Section B). When selecting the value for the penalty term λ that yields the highest AUC (Model 1),

TABLE 2 | Areas under the curves (AUCs) for subtests and results of the LASSO logistic regression models.

	Section A			Section B	
	r_{pb}	AUC	95%-CI DeLong	Lasso Model 1 – AUC = MAX	Lasso Model 2 – 1 SE rule
				Estimate (OR)	Estimate (OR)
Phonological awareness					
Rhyme detection	–0.042	0.521	(0.433, 0.610)		
Syllable count (one syllable)	–0.075	0.552	(0.467, 0.637)		
Syllable count (two or more syllables)	–0.086	0.555	(0.467, 0.643)		
Initial phoneme detection	–0.211***	0.663	(0.588, 0.738)	–0.061 (0.941)	
Rapid Automatized Naming					
RAN objects	0.287***	0.726	(0.658, 0.795)	0.312 (1.366)	0.174 (1.190)
RAN digits	0.255***	0.717	(0.632, 0.802)	0.098 (1.103)	
Letter knowledge					
Letter knowledge	–0.280***	0.723	(0.645, 0.801)	–0.380 (0.684)	–0.179 (0.836)
Phonological working memory					
Word list memory	–0.149**	0.601	(0.517, 0.686)		
Letter–number sequences forward	–0.084	0.589	(0.506, 0.671)		
Letter–number sequences backward	–0.184***	0.627	(0.541, 0.714)		
Linguistic competences					
Vocabulary	–0.184***	0.615	(0.523, 0.706)		
Sentence repetition	–0.189***	0.642	(0.556, 0.727)	–0.040 (0.961)	
Intelligence					
Subtest A	–0.072	0.581	(0.504, 0.658)		
Subtest B	–0.095	0.566	(0.477, 0.654)		
Intercept				–1.675	–1.567

** $p < 0.01$, *** $p < 0.001$.

the model identifies five non-zero predictors: initial phoneme detection ($b = -0.061$), RAN objects ($b = 0.312$), RAN digits ($b = 0.098$), letter knowledge ($b = -0.380$), and sentence repetition ($b = -0.04$). Remember that all subtests were z-scored, thus the strength of the regression coefficients could be directly compared. Applying the one standard error rule for the selection of λ (Model 2) results in two non-zero predictors. RAN objects ($b = 0.174$) and letter knowledge ($b = -0.179$)—the strongest predictors of Model 1—were selected as meaningful predictors.

The screening score based on LASSO Model 1 yields an AUC of 0.783 (DeLong 95%-CI [0.713, 0.852]), and the AUC for the screening score based on LASSO Model 2 is 0.773 (DeLong 95%-CI [0.704, 0.843]). A bootstrapped test for paired ROC curves indicates that the AUC-Difference is statistically not significant ($\Delta\text{AUC} = 0.01$, $D = 1.317$, $p = 0.188$). However, as AUC-difference tests are known to be plagued with low power [e.g., (73)], we decided to further evaluate the screening scores based on LASSO Models 1 and 2.

Table 3 reports the results for the comparison of ROC curves between groups. The tests for unpaired ROC curves show no significant differences between the groups. Thus, the screenings based on the LASSO selected predictors show no differences in diagnostic accuracy between German and non-German-speaking children, children with German-language exposure ≤ 2 years, children with German-language exposure > 2 years, and girls and boys.

Finally, we estimated cutoffs for both screening scores by setting the sensitivity equal to 0.80. This cutoff was chosen to achieve acceptable sensitivity while holding the rate of positive screens as low as possible. The diagnostic accuracy statistics are reported in **Table 4**. Both screening scores yield a sensitivity of 0.808 (i.e., the cutoff value that achieves a sensitivity closest to 0.8). For Model 1, the cutoff is 0.195. This cutoff results in 36.3% of screening fails. Notably, given that the screening

scores based on Model 1 and Model 2 achieve an identical sensitivity, the other diagnostic accuracy statistics are favoring Model 1. Importantly, as indicated by a significant McNemar Test [$\chi^2(1) = 4.923$, $p < 0.05$], the Model 1 screening turns out to be significantly more specific than the Model 2 screening (Model 1: Sp = 0.733, 95%-CI [0.672, 0.787], Model 2: Sp = 0.695, 95%-CI [0.633, 0.753]).

For Model 2, the cutoff of 0.186 results in 39.3% of screening fails. Given that the screening scores based on Model 1 and Model 2 achieve an identical sensitivity, the other diagnostic accuracy statistics of Model 1 are better than those of Model 2. Notably, as indicated by a significant McNemar Test [$\chi^2(1) = 4.923$, $p < 0.05$], the Model 1 screening turns out to be significantly more specific (Model 1: Sp = 0.733, 95%-CI [0.672, 0.787], Model 2: Sp = 0.695, 95%-CI [0.633, 0.753]).

DISCUSSION

Constructing a New Screening Battery

The analysis of a broad battery of subtests aiming to predict word reading deficiencies at the end of first grade resulted in two models: a short one consisting of two subtests (AUC = 0.77) and a broader one with five subtests (AUC = 0.78). Both models include a task for rapid naming and letter knowledge; the broader version additionally includes two language subtests (vocabulary and grammar) and a short assessment of phonological awareness (first phoneme detection).

There is the general consensus about the acceptable test accuracy of developmental screenings, namely, a sensitivity of 0.80 and specificity of 0.70 (74, 75). Thus, moderate specificity values may be acceptable, but high sensitivity is demanded for universal screening (76, 77). Setting a sensitivity range of 0.80, the specificity of the short version with two predictors is 0.69, and the

TABLE 3 | Comparing receiver operating characteristics (ROC) curves between subsamples.

		LASSO Model 1			LASSO Model 2		
		AUC	95%-CI	AUC-Difference (2)	AUC	95%-CI	AUC-Difference (2)
First language	German (1)	0.768	(0.671, 0.866)	$E = 0.038, p = 0.502$ $D = 0.248, p = 0.804$	0.761	(0.664, 0.857)	$E = 0.041, p = 0.417$ $D = 0.202, p = 0.8401$
	Non-German (2)	0.786	(0.680, 0.892)				
German language exposure	≤ 2 years (1)	0.767	(0.681, 0.856)	$E = 0.015, p = 0.780$ $D = -0.550, p = 0.582$	0.760	(0.674, 0.846)	$E = 0.014, p = 0.808$ $D = -0.385, p = 0.700$
	> 2 years (2)	0.810	(0.687, 0.934)				
Gender	Boys (1)	0.755	(0.643, 0.866)	$E = 0.017, p = 0.695$ $D = -0.754, p = 0.451$	0.744	(0.632, 0.856)	$E = 0.02, p = 0.576$ $D = -0.834, p = 0.405$
	Girls (2)	0.809	(0.721, 0.897)				

E-value for AUC-Difference refers to the Venkatraman test that compares ROC curves and the *D*-value refers to the bootstrapped test for paired ROC curves that compares AUCs.

TABLE 4 | Diagnostic accuracy statistics.

	Se (95%-CI)	Sp (95%-CI)	PPV (95%-CI)	NPV (95%-CI)	DLR+ (95%-CI)	DLR- (95%-CI)
LASSO Model 1	0.808 (0.675, 0.904)	0.733 (0.672, 0.787)	0.393 (0.326, 0.591)	0.947 (0.898, 0.960)	3.012 (2.360, 3.865)	0.263 (0.150, 0.461)
LASSO Model 2	0.808 (0.675, 0.904)	0.695 (0.633, 0.753)	0.362 (0.300, 0.559)	0.944 (0.893, 0.957)	2.652 (2.104, 3.344)	0.277 (0.157, 0.486)

specificity of the extended model with three additional subtests is 0.74. Yielding significantly higher specificity, the extended model is thus the preferred model.

The positive predictive value of Model 1 is 0.36; that is, 36% of the children with low screening results are in the slow readers' group at the end of Grade 1. The preferable five-variable model identifies 39% of the children with low reading results at the end of first grade, identifying 42 children at risk as low readers correctly. The achieved predictive values represent an improvement compared to the only recent assessment for preschoolers in German, the "LRS-Screening" (50) which uses a range of 14 subtests in the year prior to starting school to predict word reading deficits at the end of first grade, with a sensitivity of 0.74, a specificity of 0.68, and a PPV of 0.27. The "LRS-Screening" is presented in a one-to-one setting in a paper-pencil version. It lasts for a duration of about 30 min, requires additional scoring time, and does not provide a cover story that would be assumed to make the assessment more appealing. In comparison, the five-component model of the newly developed screening tool can now be administered in about 15 min, including a shortened cover story, and is the only app-based screening tool for the early identification of reading problems in German. The "Bielefelder Screening" (78) is widely used in the year prior to school entry. In the manual, good predictive values (as high as 50%) are quoted, which could not be replicated by independent studies (79). Another screening, designed for group assessment during the last year of kindergarten, is the Phonological Awareness-Reading and Spelling Screening [PB-LRS; (80)]. The authors reported a sensitivity of 63%, specificity of 87%, and PPV of 36%. The duration of this screening tool with acceptable predictive quality is about 60 min.

Another established screening tool is called "Tour through Hörhausen" (81), which provides a phonological assessment of children in one-to-one settings at the beginning and the midpoint of the first grade. Prognostic validity was analyzed using a sample of 375 children, focusing on word reading speed at first grade. The authors described its specificity as over 80%, whereas the sensitivity varies between 38 and 48%. The assessment time is about 40 min. In summary, established screenings to predict reading difficulties in German require a long administration time and demonstrate low predictive power [for an overview, see (82)].

With the five-component model, there is no significant difference in the prediction of reading deficits according to gender or first language (German or non-German). Therefore, no specific cutoffs for gender or first language are needed.

Phonological awareness is one component of phonological information processing that is highly significant for the prediction of reading deficits in the international English-dominated literature [e.g., (83) for an overview]; in more consistent orthographies, such as Italian or German, word reading deficits are primarily predicted by the measures of letter knowledge and RAN (54, 84, 85). The prediction of reading performance at word level by vocabulary and grammar, summarized as linguistic competencies (34), was confirmed in the present study for the German language.

Interestingly, the factor of non-verbal intelligence plays a subordinate role; in the statistical model, it does not attain significance. In German-language longitudinal studies, the

predictive quality of non-verbal intelligence on reading fluency was minimal (86, 87). As in the present study, factors specific to reading and writing, such as RAN and letter knowledge, showed higher predictive power for reading difficulties than the general factor of non-verbal intelligence.

Family history of reading problems was not included in our analysis, although prior studies found some contribution to a prediction model (36). However, a recent longitudinal study with a representative, epidemiological sample did not report acceptable AUC values for predicting reading problems by eliciting family risk factors (88), therefore diminishing the predictive value of family risk factors. This effect is expected especially for a German-speaking country because there are usually reservations about reporting family predispositions, and therefore no or unreliable information is provided.

School Use of the Screening Battery

For use in schools, screening tools should not only have high predictive power and reliability but also should have applicability with limited resources. In addition, screenings should be highly motivating for children. Children indicated that they experienced the assessment as a game and were able to stay with it well over a median duration of about 38 min. Not a single child had to stop for reasons of motivation or declining attention. The identification figure SCHWUPP provided continuous positive feedback after each completed subtest, and the frame story between the tasks could be used for relaxation. The new screening tool can be administered in about 15 min, making it shorter than any other screening tool available for the German language.

For the testers themselves, a high degree of objectivity was ensured because all instructions were played as audio files and important additional information (e.g., when naming the letters) was documented in the app. Due to the high degree of standardization, the training effort was low. Given that the one-to-one test setting remains necessary since some tasks require a paper target (such as letter knowledge or RAN) and the screening tool is for young children, a contact person is important in stressful situations. For teachers, rapid feedback of the results through automatic uploading of the results and further evaluations by the project team was important. Furthermore, for teachers, automatic scoring is regarded as a key feature of a feasible instrument.

Strengths of the New Screening Tool

The comprehensive sample is representative of Austria and the German language. A five-variable screening for surveillance in a community school setting showed good predictive power to detect slow readers at the end of first grade. For children, a motivating cover story presented interactively through tablets helps to maintain their motivation through a series of tasks. Advantages for screeners are short administration time, objectivity, and a quick computation of results. In order to make the whole screening tool available to primary schools without licensing costs, sub-tests had to be newly designed. There is now a screening tool that meets these requirements.

Limitations

Although a PPV of 39% is good compared to given screening tools, it does not cover all the children with slow reading at the end of first grade. Some data about environmental factors have been captured, but a big amount of variance is still to be detected: reading socialization *via* parents, school, friends, and at a macro level, society. With regard to the predictors, it must be noted that only language-related variables were included. Evidence on preschool visual processing has also recently been shown to be predictive of the reading process. Visual predictors were unfortunately not collected in the present study. Finally, continuous surveillance of reading is required because there might be different pathways to reading difficulties (many children with early difficulties do not develop later reading problems, and many children who do not fail the initial screening demonstrate reading difficulties later on).

Implications for Research and Practice

The first steps for a new screening on reading deficits have been implemented. Further validation of the newly constructed screening is needed, the next steps include a bigger normative sample and comparisons with screening tools already in use. Feasibility data for school usage must be gathered from children and teachers in order to enhance the screening and support a broader and well-accepted rollout.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

REFERENCES

- Ehri LC. Phases of development in learning to read words by sight. *J Res Read.* (1995) 18:116–25. doi: 10.1111/j.1467-9817.1995.tb00077.x
- Stanovich K, West R. *Read Res Quart.* (1989) 24:747605. doi: 10.2307/747605
- Esser G, Wyschkon A, Schmidt M. Was wird aus achtjährigen mit einer lese- und rechtschreibstörung: ergebnisse im alter von 25 jahren. *Zeitschrift Für Klinische Psychologie und Psychotherapie Forschung Praxis.* (2002) 31:235–42. doi: 10.1026//1616-3443.31.4.235
- Gasteiger-Klicpera B, Klicpera C, Schabmann A. Der Zusammenhang zwischen Lese-, Rechtschreib- und Verhaltensschwierigkeiten. *Kindheit und Entwicklung.* (2006) 15:55–67. doi: 10.1026/0942-5403.15.1.55
- Kohn J, Wyschkon A, Ballaschk K, Ihle W, Esser G. Verlauf von umschriebenen entwicklungsstörungen: eine 30-monats-follow-up-studie. *Lernen Lernstörungen.* (2013) 2:77–89. doi: 10.1024/2235-0977/a000032
- Landerl K. Lese-/Rechtschreibstörung. In: Schneider S, Margraf J editors. *Lehrbuch der Verhaltenstherapie: Band 3: Störungen im Kindes- und Jugendalter.* Berlin: Springer (2009). p. 395–410. doi: 10.1007/978-3-540-79545-2_25
- Moll K, Hutzler F, Wimmer H. Developmental dyslexia in a regular orthography: a single case study. *Neurocase.* (2005) 11:433–40. doi: 10.1080/13554790500263537
- Moll K, Kunze S, Neuhoff N, Bruder J, Schulte-Körne G. Specific learning disorder: prevalence and gender differences. *PLoS One.* (2014) 9:e103537. doi: 10.1371/journal.pone.0103537
- Schulte-Körne G. Diagnostik und Therapie der Lese-Rechtschreib-Störung. *Deutsches Ärzteblatt Int.* (2010) 107:718–27. doi: 10.3238/arztebl.2010.0718
- Roth E, Warnke A. Diagnose und Therapie der Lese-Rechtschreibstörung. In: Koletzko B, Reinhardt D, Stöckler-Ipsiroglu S editors. *Pädiatrie upgrade 2002: Weiter- und Fortbildung.* Berlin: Springer (2002). p. 166–75. doi: 10.1007/978-3-642-56123-8_17
- Perfetti CA. The representation problem in reading acquisition. In: Gough PB, Ehri LC, Treiman R editors. *Reading Acquisition.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc (1992). p. 145–74. doi: 10.4324/9781351236904-6
- Perfetti CA, Hart L. The lexical bases of comprehension skill. In: Gorfien D editor. *On the Consequences of Meaning Selection.* Washington, DC: American Psychological Association (2001). doi: 10.1037/10459-004
- McElvany N, Schneider C. Förderung von Lesekompetenz. In: Lenhard W, Schneider W editors. *Diagnostik und Förderung des Leseverständnisses.* Göttingen: Hogrefe (2009). p. 151–84.
- Duncan GJ, Dowsett CJ, Claessens A, Magnuson K, Huston AC, Klebanov P, et al. School readiness and later achievement. *Dev Psychol.* (2007) 43:1428–46. doi: 10.1037/0012-1649.43.6.1428
- Jordan N, Kaplan D, Locuniak M, Ramineni C. Predicting First-Grade Math Achievement from Developmental Number Sense Trajectories. *Learn Disabil Res Pract.* (2007) 22:36–46. doi: 10.1111/j.1540-5826.2007.00229.x
- Moffitt T, Caspi A, Harrington H, Milne B. Males on the life-course-persistent and adolescence-limited antisocial pathways: follow-up at age 26 years. *Dev Psychopathol.* (2002) 14:179–207. doi: 10.1017/S0954579402001104

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Regional School Board for Upper Austria (Bildungsdirektion Upper Austria). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

MS and GS: conceptualization, investigation, data curation, and project administration. MS, GS, and CW: methodology. CW: formal analysis. MS, GS, CW, and DH: writing original draft preparation and reviewing and editing. All authors contributed to the article and approved the submitted version.

FUNDING

All costs related to this research project were borne by the University of Education Upper Austria and Research Institute for Developmental Medicine (RID), Johannes Kepler University Linz. The tablets were provided by the project team with funds from the University of Education Upper Austria. Article processing charge was funded by the Johannes Kepler University Linz Open Access Publishing Fund.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fped.2022.863477/full#supplementary-material>

17. Nordström T, Jacobson C, Söderberg Juhlander P. Early word decoding ability as a longitudinal predictor of academic performance. *Eur J Psychol Educat.* (2015) 2015:1–17. doi: 10.1007/s10212-015-0258-5
18. McLaughlin MJ, Speirs KE, Shenassa ED. Reading disability and adult attained education and income: a 30-year longitudinal study of a population-based sample. *J Learn Disabil.* (2014) 47:374–86. doi: 10.1177/0022219412458323
19. Klicpera C, Gasteiger-Klicpera B, Schabmann A. *Lesen und Schreiben—Entwicklung und Schwierigkeiten. Die Wiener La?ngsschnittuntersuchungen u?ber die Entwicklung, den Verlauf und die Ursachen von Lese- und Schreibschwierigkeiten in der Pflichtschulzeit.* Bern: Huber (1993).
20. Moll K, Gangl M, Banfi C, Schulte-Körne G, Landerl K. Stability of Deficits in Reading Fluency and/or Spelling. *Sci Stud Read.* (2020) 24:241–51. doi: 10.1080/10888438.2019.1659277
21. Livingston EM, Siegel LS, Ribary U. Developmental dyslexia: emotional impact and consequences. *Austral J Learn Difficult.* (2018) 23:107–35. doi: 10.1080/19404158.2018.1479975
22. Fischbach A, Schuchardt K, Mähler C, Hasselhorn M. Zeigen Kinder mit schulischen Minderleistungen sozio-emotionale Auffälligkeiten? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychol.* (2010) 42:201–10. doi: 10.1026/0049-8637/a000025
23. Mammarella IC, Ghisi M, Bomba M, Bottesi G, Caviola S, Broggi F, et al. Anxiety and depression in children with nonverbal learning disabilities, reading disabilities, or typical development. *J Learn Disabil.* (2016) 49:130–9. doi: 10.1177/0022219414529336
24. Galuschka K, Von Wirth E, Krick K, Schulte-Körne G. Effectiveness of treatment approaches for children and adolescents with reading disabilities: a meta-analysis of randomized controlled trials. *PLoS One.* (2014) 9:e89900. doi: 10.1371/journal.pone.0089900
25. Amelang M, Schmidt-Atzert L. *Psychologische Diagnostik und Intervention.* Berlin: Springer (2006).
26. Paleczek L, Seifert S, Schöfl M. Comparing digital to print assessment of receptive vocabulary with GraWo-KiGa in Austrian kindergarten. *Br J Educat Technol.* (2021) 52:13163. doi: 10.1111/bjet.13163
27. Thompson P, Hulme C, Nash H, Gooch D, Hayiou-Thomas E, Snowling M. Developmental dyslexia: predicting individual risk. *J Child Psychol Psychiat Allied Discipl.* (2015) 56:12412. doi: 10.1111/jcpp.12412
28. Grainger J, Dufau S, Ziegler JC. A Vision of Reading. *Trends Cognit Sci.* (2016) 20:171–9.
29. Vernet M, Bellocchi S, Leibnitz L, Chaix Y, Ducrot S. Predicting future poor readers from pre-reading visual skills: a longitudinal study. *Appl Neuropsychol Child.* (2021) 2021:1895790. doi: 10.1080/21622965.2021.1895790
30. Landerl K, Castles A, Parrila R. Cognitive precursors of reading: a cross-linguistic perspective. *Sci Stud Read.* (2021) 2021:1983820. doi: 10.1080/10888438.2021.1983820
31. Ptok M, Berendes K, Gottal S, Grabherr B, Schneeberg J, Wittler M. Phonologische Verarbeitung. *Monatsschrift Kinderheilkunde.* (2008) 156:860–6. doi: 10.1007/s00112-008-1770-3
32. Schnitzler CD, Springer L, Schrey-Dern D. *Phonologische Bewusstheit und Schriftspracherwerb.* Stuttgart: Thieme Verlag (2008). doi: 10.1055/b-002-21540
33. Pinto G, Bigozzi L, Tarchi C, Accorti Gamannossi B, Canneti L. Cross-lag analysis of longitudinal associations between primary school students' writing and reading skills. *Read Writ.* (2015) 28:9. doi: 10.1007/s11145-015-9569-9
34. Snowling M, Moll K, Hulme C. Language difficulties are a shared risk factor for both reading disorder and mathematics disorder. *J Exp Child Psychol.* (2021) 202:105009. doi: 10.1016/j.jecp.2020.105009
35. Price KM, Wigg KG, Misener VL, Clarke A, Yeung N, Blokland K, et al. Language Difficulties in School-Aged Children With Developmental Dyslexia. *J Learn Disabil.* (2021) 2021:2221942110062. doi: 10.1177/00222194211006207
36. Ozernov-Palchik O, Norton E, Sideridis G, Beach S, Wolf M, Gabrieli J, et al. Longitudinal stability of pre-reading skill profiles of kindergarten children: implications for early screening and theories of reading. *Dev Sci.* (2016) 20:12471. doi: 10.1111/desc.12471
37. Hurrelmann B, Becker S, Nickel-Bacon I. *Lesekindheiten. Familie und Lesesozialisation im historischen Wandel.* Weinheim: Juventa (2006).
38. Rosebrock C, Nix D. *Grundlagen der Lesedidaktik und der systematischen schulischen Leseförderung.* Hohengehren: Baltmannsweiler: Schneider-Verl (2008).
39. Rosebrock C, Nix D, Rieckmann C, Gold A. *Leseflüssigkeit fördern. Lautleseverfahren für die Primar- und Sekundarstufe.* Seelze: Kallmeyer (2011).
40. Ibáñez Alfonso JA, Hernández Cabrera JA, Duñabeitia JA, Monzó A, Macizo Soria P, Bajo Molina MT, et al. Socioeconomic status, culture, and reading comprehension in immigrant students. *Front Psychol.* (2021) 12:752273. doi: 10.3389/fpsyg.2021.752273
41. Lervåg A, Aukrust VG. Vocabulary knowledge is a critical determinant of the difference in reading comprehension growth between first and second language learners. *J Child Psychol Psychiat.* (2010) 51:612–20. doi: 10.1111/j.1469-7610.2009.02185.x
42. Leikin M, Schwartz M, Share DL. General and specific benefits of bi-literate bilingualism: a Russian–Hebrew study of beginning literacy. *Read Writ.* (2010) 23:269–92. doi: 10.1007/s11145-009-9210-x
43. Lipka O, Siegel LS. The development of reading comprehension skills in children learning English as a second language. *Read Writ.* (2012) 25:1873–98. doi: 10.1007/s11145-011-9309-8
44. Shany M, Geva E. Cognitive, language, and literacy development in socio-culturally vulnerable school children – the case of Ethiopian Israeli children. In: Leikin M, Schwartz M, Tobin Y editors. *Current Issues in Bilingualism.* Dordrecht: Springer Netherlands (2012). p. 77–117. doi: 10.1007/978-94-007-2327-6_5
45. Proctor CP, Silverman RD, Harring JR, Montecillo C. The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish–English bilingual children in elementary school. *Read Writ.* (2012) 25:1635–64. doi: 10.1007/s11145-011-9336-5
46. Geva E, Zadeh Z. Reading efficiency in native english-speaking and english-as-a-second-language children: the role of oral proficiency and underlying cognitive-linguistic processes. *Sci Stud Read.* (2006) 10:31–57. doi: 10.1207/s1532799xssr1001_3
47. Farnia F, Geva E. Growth and predictors of change in English language learners' reading comprehension. *J Res Read.* (2013) 36:389–421. doi: 10.1111/jrir.12003
48. Duzy D, Ehm J-H, Souvignier E, Schneider W, Gold A. Prädiktoren der Lesekompetenz bei Kindern mit Deutsch als Zweitsprache. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychol.* (2013) 45:173–90. doi: 10.1026/0049-8637/a000093
49. Statistik Austria. *Bildung in Zahlen. 2019/20. Schlüsselindikatoren und Analysen.* Wien: Statistik Austria (2021).
50. Endlich D, Lenhard W, Marx P, Scheider W. *LRS-Screening Laute, Reime, Sprache - Würzburger Screening zur Früherkennung von Leseschreibschwierigkeiten.* Göttingen: Hogrefe (2019).
51. Schroeder S, Würzner KM, Heister J, Geyken A, Kliegl R. childLex—eine lexikalische datenbank zur schriftsprache für kinder im Deutschen. *Psychol Rundsch.* (2015) 66:155–65. doi: 10.1026/0033-3042/a000275
52. Muthén LK, Muthén BO. *Mplus user's guide. Eighth edition.* Los Angeles, CA: Muthén & Muthén (1998-2017).
53. Denckla MB, Rudel R. Rapid “automatized” naming of pictured objects, colors, letters and numbers by normal children. *Cortex J Devoted Study Nerv Syst Behav.* (1974) 10:186–202. doi: 10.1016/S0010-9452(74)80009-2
54. Landerl K, Ramus F, Moll K, Lyytinen H, Leppänen P, Lohvansuu K, et al. Predictors of developmental dyslexia in European orthographies with varying complexity. *J Child Psychol Psychiat Allied Discipl.* (2012) 54:12029. doi: 10.1111/jcpp.12029
55. Grob A, Meyer CS, Hagemann-von Arx P. *Intelligence and Development Scales (IDS). Intelligenz- und Entwicklungsskalen für Kinder von 5-10 Jahren.* Bern: Verlag Hans Huber (2009).
56. Marinis T, Armon-Lotem S. *Sentence Repetition.* Bristol: Multilingual Matters (2015). p. 95–124. Available online at: <https://kops.uni-konstanz.de/handle/123456789/42352>.
57. Ibrahim L, Hamann C, Öwerdieck D. *Identifying Specific Language Impairment (SLI) across Different Bilingual Populations: German Sentence Repetition Task (SRT).* (2018). Available

- online at: [https://www.semanticscholar.org/paper/Identifying-Specific-Language-Impairment-\(-SLI-\)-%3A-Ibrahim-Hamann/48cf396157502fa5835f8bc09c180086a30945ac](https://www.semanticscholar.org/paper/Identifying-Specific-Language-Impairment-(-SLI-)-%3A-Ibrahim-Hamann/48cf396157502fa5835f8bc09c180086a30945ac) (accessed November 24, 2021).
58. Seifert S, Paleczek L, Schwab S, Gasteiger-Klicpera B. *GraWo: Grazer Wortschatztest*. Göttingen: Hogrefe (2017).
 59. Wyschkon A, Esser G. *PITVA: Potsdamer Intelligenztest für das Vorschulalter: Manual*. Göttingen: Hogrefe (2019).
 60. Lenhard W, Lenhard A, Schneider W. *Ein Leseverständnistest für Erst- bis Siebtklässler (ELFE II)*. Göttingen: Hogrefe (2017).
 61. Swets JA. Measuring the accuracy of diagnostic systems. *Science*. (1988) 240:1285–93. doi: 10.1126/science.3287615
 62. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform*. (2011) 12:77. doi: 10.1186/1471-2105-12-77
 63. McNeish D. Using lasso for predictor selection and to assuage overfitting: a method long overlooked in behavioral sciences. *Multivar Behav Res*. (2015) 50:471–84. doi: 10.1080/00273171.2015.1036965
 64. Gareth J, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*. Heidelberg: Springer (2013).
 65. Silverman RD, McNeish D, Speece DL, Ritchey KD. Early screening for decoding- and language-related reading difficulties in first and third grades. *Assess Effect Interv*. (2021) 46:99–109.
 66. Friedman J, Hastie T, Tibshirani R, Narasimhan B, Tay K, Simon N, et al. *Package 'glmnet'*. CRAN R Repository. Vienna: R Core Team (2021).
 67. Youngstrom EA. A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: we are ready to ROC. *J Pediatr Psychol*. (2014) 39:204–21. doi: 10.1093/jpepsy/jst062
 68. Venkatraman ES. A Permutation Test to Compare Receiver Operating Characteristic Curves. *Biometrics*. (2000) 56:1134–8. doi: 10.1111/j.0006-341x.2000.01134.x
 69. López-Ratóns M, Rodríguez-Álvarez M, Cadarso-Suárez C, Gude F. OptimalCutpoints: an R package for selecting optimal cutpoints in diagnostic tests. *J Statist Softw*. (2014) 61:i08. doi: 10.18637/jss.v061.i08
 70. Greenslade K, Plante E, Vance R. The diagnostic accuracy and construct validity of the structured photographic expressive language test—preschool: second edition. *Lang Speech Hearing Serv Sch*. (2008) 40:150–60. doi: 10.1044/0161-1461(2008)07-0049
 71. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford: University Press (2003).
 72. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature: III. How to use an article about a diagnostic test: A. Are the results of the study valid? *J Am Med Assoc*. (1994) 271:389–91. doi: 10.1001/jama.1994.03510290071040
 73. Seshan VE, Gönen M, Begg CB. Comparing ROC curves derived from regression models. *Statist Med*. (2013) 32:1483–93. doi: 10.1002/sim.5648
 74. Catts H, Nielsen D, Bridges M, Liu Y-S. Early Identification of Reading Comprehension Difficulties. *J Learn Disabil*. (2014) 49:22219414556121. doi: 10.1177/0022219414556121
 75. Kilgus S, Eklund K. Consideration of base rates within universal screening for behavioral and emotional risk: a novel procedural framework. *School Psychol Forum*. (2016) 10:120–30.
 76. Compton DL, Fuchs D, Fuchs LS, Bryant JD. Selecting at-risk readers in first grade for early intervention: a two-year longitudinal study of decision rules and procedures. *J Educat Psychol*. (2006) 98:394–409. doi: 10.1037/0022-0663.98.2.394
 77. Jenkins J, Hudson R, Johnson E. Screening for At-Risk Readers in a Response to Intervention Framework. *School Psychol Rev*. (2007) 36:12087919. doi: 10.1080/02796015.2007.12087919
 78. Jansen H, Mannhaupt G, Marx H, Skowronek H. *Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten (BISC)*. Göttingen: Hogrefe (1999).
 79. Marx P, Weber J. Vorschulische Vorhersage von Lese- und Rechtschreibschwierigkeiten. *Zeitschrift Für Pädagogische Psychol*. (2006) 20:251–9. doi: 10.1024/1010-0652.20.4.251
 80. Barth K, Gomm B. *Gruppentest zur Früherkennung von Lese- und Rechtschreibschwierigkeiten (PB-LRS)*. München: Reinhardt (2004).
 81. Martschinke S, Kirschhock EM, Frank A. *Diagnose und Förderung im Schriftspracherwerb. Band 1. Der Rundgang durch Hörhausen. Erhebungsverfahren zur phonologischen Bewusstheit*. Hamburg: Auer-Verlag (2001).
 82. Marx P, Lenhard W. Diagnostische Merkmale von Screeningverfahren. In: Hasselhorn M, Schneider W editors. *Fru?hprognose schulischer Kompetenzen*. Göttingen: Hogrefe (2010).
 83. Rose J. *Identifying and teaching children and young people with dyslexia and literacy difficulties*. Nottingham: DCSF (2009).
 84. Lohvansuu K, Torppa M, Ahonen T, Eklund K, Hämäläinen J, Leppanen P, et al. Unveiling the Mysteries of Dyslexia—Lessons Learned from the Prospective Jyväskylä Longitudinal Study of Dyslexia. *Brain Sci*. (2021) 11:427. doi: 10.3390/brainsci11040427
 85. Huschka S, Georgiou G, Brandenburg J, Ehm J-H, Hasselhorn M. Examining the Contribution of RAN Components to Reading Fluency, Reading Comprehension, and Spelling in German. *Read Writ*. (2021) 34:2317–36. doi: 10.1007/s11145-021-10145-6
 86. Ennemoser M, Marx P, Weber J, Schneider W. Spezifische vorläuferfertigkeiten der lesegeschwindigkeit, des leseverständnisses und des rechtschreibens: evidenz aus zwei längsschnittstudien vom kindergarten bis zur 4. klasse. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychol*. (2012) 44:53–67. doi: 10.1026/0049-8637/a000057
 87. Landerl K, Wimmer H. Development of Word Reading Fluency and Spelling in a Consistent Orthography: an 8-Year Follow-Up. *J Educat Psychol*. (2008) 100:150–61. doi: 10.1037/0022-0663.100.1.150
 88. Ferrer E, Shaywitz B, Holahan J, Shaywitz S. Family History Is Not Useful in Screening Children for Dyslexia. *J Pediatr Neuropsychol*. (2021) 8:15–21. doi: 10.1007/s40817-021-00110-0
 89. Stanovich K. Matthew effects in reading: some consequences of individual differences in the acquisition of literacy. *Read Res Q*. (1986) 21:360–407. doi: 10.1598/RRQ.21.4.1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Schöfl, Steinmair, Holzinger and Weber. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.