

Ana Bassôa¹ 
Adriana C. Costa² 
Rudineia Toazza² 
Augusto Buchweitz³ 

Keywords

Screening
Dyslexia
School Teachers
Psychometrics
Validation Study

Descritores

Rastreio
Dislexia
Professores Escolares
Psicometria
Estudo de Validação

Correspondence address:

Augusto Buchweitz
Escola de Ciências da Saúde e da Vida,
Pontifícia Universidade Católica do
Rio Grande do Sul – PUCRS
Av. Ipiranga, 6690, Jardim Botânico,
Prédio 63, Porto Alegre (RS), Brasil,
CEP: 90610-000.
E-mail: augusto.buchweitz@puers.br

Received: April 02, 2020

Accepted: September 06, 2020

Scale for developmental dyslexia screening: evidence of validity and reliability

Escala para rastreamento de dislexia do desenvolvimento: evidências de validade e fidedignidade

ABSTRACT

Purpose: To investigate the empirical validity and reliability of a screener for risk of developmental dyslexia (DD) by elementary school teachers. **Methods:** The scale was tested with 12 teachers who answered questions about their students (95 students total, all in the third year of elementary school); the students, in turn, performed reading and writing tasks which were used to investigate the association between screening scores and performance. The following analyses were carried out: (1) factor analysis; (2) internal consistency; (3) relationship between each scale item and the construct of interest, as measured by item response theory (IRT); (4) correlation of each scale item with external variables (reading and writing tests); and (5) the temporal stability of teachers' evaluations. **Results:** The analyses showed: (1) one factor was extracted; (2) strong internal consistency – the items in the scale are good indicators for screening of this construct; (3) items were monotonic (IRT), i.e., item variability is associated with one construct; (4) moderate Spearman correlation (11/17 items); (5) temporal stability – the result of screening did not vary over time. **Conclusion:** This study shows evidence of validity and reliability of the proposed scale in its intended use of screening for developmental dyslexia. The percentage of children at risk for developmental dyslexia, according to the scale, was approximately 9%, which is in agreement with the international literature on the prevalence of dyslexia.

RESUMO

Objetivo: Investigar validade e fidedignidade de uma escala de rastreamento para dislexia do desenvolvimento (DD) no ensino fundamental preenchida por professores. **Método:** Avaliação empírica - 12 professores responderam a Escala de Leitura e Escrita (ELE) sobre 95 alunos de 3º ano do ensino fundamental, em dois momentos; os escolares realizaram testes de leitura e escrita (variáveis externas) para investigar a correlação entre a escala e o desempenho dos mesmos. Realizaram-se (1) análise fatorial, (2) avaliação da consistência interna, (3) investigação da relação entre um item da escala e o construto medido por teoria da resposta ao item (TRI) (4) correlação da escala com variáveis externas (Validade Convergente-VC); e (5) investigação da estabilidade temporal da avaliação. **Resultados:** (1) a escala avalia um único fator; (2) o coeficiente alpha apontou que os itens são bons indicadores do construto; (3) a análise por TRI mostra que todos os itens foram monotônicos, indicando que um único construto determina a variabilidade (4) a correlação de Spearman foi moderada (11/17 itens), apontando a existência de VC; (5) o valor da correlação do coeficiente de estabilidade temporal indica que o resultado da ELE não varia de maneira significativa no tempo; (6) nove crianças obtiveram pontuações que sugerem encaminhamento para uma avaliação diagnóstica devido ao grau de dificuldade apresentado. **Conclusão:** O estudo mostra evidências empíricas de validade e fidedignidade da ELE para rastreamento de risco de DD. A porcentagem de crianças com suspeita de DD (aproximadamente 9%) corrobora a literatura internacional sobre prevalência de dislexia.

Study conducted at Pontifícia Universidade Católica do Rio Grande do Sul – PUCRS - Porto Alegre (RS), Brasil.

¹ Programa de Pós-graduação em Medicina e Ciências da Saúde, Neurociências, Pontifícia Universidade Católica do Rio Grande do Sul – PUCRS - Porto Alegre (RS), Brasil.

² Instituto do Cérebro do Rio Grande do Sul – InsCer - Porto Alegre (RS), Brasil.

³ Escola de Ciências da Saúde e da Vida, Pontifícia Universidade Católica do Rio Grande do Sul – PUCRS - Porto Alegre (RS), Brasil.

Financial support: Coordenação de Aperfeiçoamento de Pessoal Nível Superior – Brasil (CAPES) – Finance code 001 - Process number: 23038.002530/2-13-93.

Conflict of interests: nothing to declare.



This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

The aim of this study was to obtain evidence of validity and reliability of a scale designed to screen for symptoms of developmental dyslexia (DD), by means of empirical analysis. This self-report scale was developed with the aim of being accessible, easy to use, and easy to analyze for teachers and other professionals who work with children during literacy development. Considering the lack of instruments for DD screening in Brazilian Portuguese, it was designed to have empirical validity and utility to aid in the identification of red flags for this learning disorder.

DD is a specific learning disorder of neurobiological origin. It is defined by an unexpected difficulty in reading for the child's chronological age, intellectual quotient, and educational level, which cannot be explained by another diagnosis⁽¹⁾. Although the DSM-5⁽¹⁾ uses the term dyslexia only as a descriptor for Specific Learning Disorder with Impairment of Reading, the term *developmental dyslexia* was retained for the present study because of its widespread and historical use.

Worldwide, dyslexia is estimated to affect 5% to 10% of readers. In Brazil, a prevalence of 7% would correspond approximately 3 million dyslexics among the 49 million students in basic education as of 2014⁽²⁾.

DD is widely underdiagnosed or diagnosed late in Brazil. A recent survey of Brazilian children with dyslexia identified that 60% of them had been held back at least once and that the average age at diagnosis was between 10 and 11 years, which suggests misinformation and a lack of screening for early diagnosis⁽³⁾. These children had already completed between 5 and 4 years of schooling without so much as being identified as at risk of DD, even though some red flags for this disorder are already manifest during preschool or first grade, as described in the literature⁽⁴⁾.

Early identification of DD mitigates student absenteeism and other harmful effects of the low reading level associated with this disorder. Easy-to-use screeners for DD red flags, such as the Screener for Reading and Writing (SRW) proposed herein, can help identify children at risk. Two points that bear stressing are the role of the elementary school teacher and the wide range of potential learning difficulties in the public school system. Studies have demonstrated the reliability of teachers' capacity to judge the reading skills of their students⁽⁵⁾. Assessment of a child's performance as compared to that of her peers by a teacher, especially with the help of a structured instrument, can be an important strategy to address the problem of early identification of children at risk of learning disorders.

The SRW can identify children who exhibit certain symptoms and behaviors characteristic of DD. It must be noted that the SRW does not replace clinical investigation for the diagnosis of DD. The SRW was based on the structure of the SNAP-IV (Swanson, Nolan, and Pelham Rating Scale) scale used to screen for attention deficit/hyperactivity disorder⁽⁶⁾ and on a theoretical review of the literature, focusing mainly on diagnostic manuals. After this first stage of development, the SRW was submitted to a panel of expert judges from different regions of the country for analysis. This analysis led to the exclusion and inclusion

of items based on the experts' assessment, as well as changes in the wording of items to ensure intelligibility and clarity. The resulting draft was submitted to elementary school teachers for semantic analysis, to ensure that there were no distorted interpretations of the items. These steps of the instrument development process are explained elsewhere⁽⁷⁾.

The choice to validate the scale at the end of the third grade was based on the DSM-5⁽¹⁾. Under Criteria C and D for Specific Learning Disorders, the DSM establishes that difficulties must arise during school years and cannot be a result of lack of educational opportunities. Considering these criteria and the National Common Core Curriculum (*Base Nacional Comum Curricular*, BNCC), which mandates that literacy must be acquired by the end of the third grade of elementary school (e.g., at which time the National Literacy Assessment is carried out), our understanding was that only from this stage onward could the diagnostic hypothesis of dyslexia be established more accurately. The third grade is a milestone of the learning process in several guidelines. It is during this year that reading difficulties are most likely to be recognized by teachers; in Brazil, it is also the first grade which a student can fail. It should be noted that the BNCC's prescriptive stance on "a certain age" is a guideline based on evidence about neurodevelopment and the optimal age for acquisition of literacy^(8,9); therefore, however arbitrary, this guideline establishes a time frame within the Brazilian educational process during which a screener instrument can assist in identification of DD risk.

The SRW is unique in Brazil. Only one other scale designed to monitor aspects of socio-emotional development such as social skills, behavioral problems, and academic skills has been validated in the country: the Social Skills Rating System (SSRS), for children aged 6 to 13⁽¹⁰⁾. To date, there is no equivalent scale to screen for signs of DD. In addition to its original nature, this scale thus addresses an unmet need for a DD screener for the Brazilian population. Making this scale available for use across the country could consolidate it as an effective, user-friendly screening instrument.

EMPIRICAL TESTING: INTERNAL CONSISTENCY, FACTOR ANALYSIS, ITEM RESPONSE THEORY, CONVERGENT VALIDITY, AND TEMPORAL STABILITY

The validity and reliability of a test can be calculated through pre-established methods⁽¹¹⁾ and subsequently used in validation of the instrument⁽¹²⁾. The selection and construction of the SRW items, as well as other evidence of validity, are described elsewhere⁽⁷⁾. The present study is limited to presenting the results of statistical evaluations based on empirical data and discussing the implications thereof.

The following analyses of validity were conducted: a) internal consistency: tests whether the variability presented by each item/task has a strong relationship with the variability of the other items, as well as with variance of the final score; b) factor analysis: tests how many behaviors the scale and its items evaluate, and the extent to which each item is a good representation of the behavior which it is intended to measure;

c) item response theory: tests the ability of each item on the scale to measure the degree of ability, or skill, presented by the respondent; d) convergent validity: tests for correlation between variation in performance in tests that measure the desired skill (external variables) and variation in SRW scores; and e) temporal stability coefficient: represents the stability of measurement over time, thus estimating the measurement error of the respondent. All of these tests were performed; a detailed presentation of the methods employed is available elsewhere⁽⁷⁾. This paper presents the results that underlie our empirical validation of the instrument and a practical discussion of the SRW items and the signs and symptoms for which it screens.

METHODS

This study was approved by the Research Ethics Committee of the Pontificia Universidade Católica do Rio Grande do Sul (ethical approval number 51215715.8.0000.5336). All teachers who participated and the parents or guardians of the evaluated students provided written informed consent.

Participants

The sample of children for this study was derived from a larger umbrella project (ethical approval numbers 30895614.5.0000.5336 and 13629513.0.0000.5336). Assessment of children with the SRW was performed by 12 elementary-school teachers (Table 1) who taught third grade at the six public schools attended by the students involved in the project. These schools serve as a convenience sample for the umbrella project. Overall, the 12 teachers evaluated 122 children with the SRW. Twenty-seven of these assessments were excluded because they did not meet the inclusion criteria: a) child > 25th percentile on Raven's (n = 13); b) no intellectual quotient evaluation (n = 6); c) incomplete reading and writing assessment (n = 7); and d) screener not completed by the teacher (n = 1). Therefore, the final sample consisted of 95 students (mean age 9.27 years, SD = 0.39; 52.6% female). The study was conducted during the 2015 school year. One month after the first assessment, all teachers were invited to participate in the retest stage. The teachers' response time ranged from 2 to 4 months. Ultimately, six teachers agreed to take part in the retest stage. At this stage, 30 children (mean age 9.25 years, SD = 0.39) were selected at random and reevaluated.

Instruments and procedures

The Screener for Reading and Writing (SRW) is a self-report instrument (see attached file). A four-point Likert scale is used to measure the frequency with which symptoms (listed in 16 items) manifest. The scale was delivered to the teachers with a list identifying the selected students. Teachers were instructed to respond within 15 days. The scale was scored as follows: each item marked "never" was assigned one point; "rarely",

two points; "sometimes", three points; and "often/always", four points. The minimum total score, which denotes no difficulty, is 16 points; the maximum score, which denotes great difficulty, is 64 points.

Reading and writing performance was assessed by means of tasks performed in groups and individually. The tasks were: a) Reading aloud of words and pseudowords⁽¹³⁾; b) Evaluation of reading comprehension of expository texts⁽¹⁴⁾; c) Dictation⁽¹⁵⁾; and) writing fluency assessment⁽⁷⁾.

Data analysis

Exploratory factor analysis was used to investigate the dimensional structure of the SRW. In this analysis, robust weighted least squares estimation was performed on a variance/covariance matrix of data from the scale in order to: 1) obtain representative results of the general population, i.e., extrapolate sample data to the general population; 2) impute missing values and ordinal data, since usual estimators such as the maximum likelihood method assume items or variables as interval data and normal distribution of these indicators⁽¹⁶⁾. In Likert-type scales, however, items are ordinal variables; in these cases, factor analyses using RWLS-type estimators tend to estimate the number of factors underlying the data more accurately and produce more consistent parametric estimates of factor loadings and correlations between factors⁽¹⁷⁾. The following model fit indices were considered: Comparative Fit Index (CFI > 0.90), Tucker-Lewis index (TLI > 0.90), Standardized Root Mean Residual (SRMR), and RMSEA (Root Mean Square Error of Approximation), all with optimal (reference) values near or < 0.08. Other indices, such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), were used to compare different hypotheses of theoretical models, with lower values indicating better fit. The reliability of each factor was estimated using Cronbach's alpha coefficient, with values > 0.6 being considered adequate⁽¹⁸⁾.

In order to assess the psychometric properties of each SRW item by the item response theory (IRT), Masters' model for partial credit scoring⁽¹⁹⁾ (1978, an extension of the Rasch dichotomous model [1960] for polytomous items) was employed. The partial credit model jointly estimates the respondent's skill level and the difficulty of the items, insofar as both parameters are represented on the same simple linear continuum, in log-odds units (*logits*); as the items and the estimates of the latent trait are measured by the same metric, the estimate of the respondent's skill will correspond to a likelihood of answer or endorsement of the item category. Its assumptions were tested by: 1) factor analysis to confirm the one-dimensional nature of the instrument; 2) monotonicity (principle by which the likelihood of endorsing a particular item category increases as the participant's latent variable increases); and 3) local independence (the test items

Table 1. Sociodemographic profile of teachers

Age (years)			Years teaching			Educational attainment (%)		
Mean (SD)	Max.	Min.	Mean (SD)	Max.	Min.	Undergrad., some	Undergrad.	Graduate
41.89 (8.44)	56	27	9.62 (7.36)	20	2	8.3	41.6	50

SD = standard deviation

are independent of each other, or do not influence each other in leading to the answer).

IRT was also used to compute infit and outfit statistics for the items. For this study, values ranging from 0.5 to 1.5 were considered; according to Wright and Linacre⁽²⁰⁾, these infit and outfit limits provide productive measurement parameters. Statistical significance (by the chi-square test) was used as a tiebreaker criterion for the two goodness-of-fit measures. For example, if any item presented unsatisfactory infit and satisfactory outfit values, comparison of the difference between the model-predicted values of each item and the actual empirical values collected was performed using the chi-square test.

Spearman correlations were calculated to analyze convergent validity. Finally, temporal stability was analyzed by means of test-retest of the teachers' responses to the scale after a 2-to-4-month interval. For an instrument to be considered reliable and temporally stable, these correlation coefficients must be equal to or greater than 0.8⁽¹²⁾. Analyses were conducted in the Statistical Package for the Social Sciences (SPSS), Mplus⁽²¹⁾ and R⁽²²⁾ software environments, including the *psych*⁽²³⁾ and *mirt*⁽²⁴⁾ packages for R.

RESULTS

Given the many different analyses conducted (factor analysis, IRT, convergent validity, test-retest), the results will be presented in separate sections below.

Factor analysis: the SRW is a single-factor instrument

Factor analysis indicated that only one factor was extracted from the scale (RLWS: $\chi^2=-6150.32$; N par=16; CFI=0.98; TLI=0.98; RMSEA=0.12; SRMR=0.05). Additional analyses (scree plot, Cattell's test⁽²⁵⁾, Horn's parallel analyses⁽²⁶⁾) provided further evidence of a single factor. The factor matrix was indicative

of excellent factor loadings (0.72 to 0.96) and item covariance (Supplementary Table 01 of the supplementary material).

Analysis of internal consistency (Supplementary Table 02 of the supplementary material) showed high Cronbach's alpha coefficients (0.968), indicating a high degree of covariance between the items on the scale (the lowest alpha was 0.9643, and the highest, 0.9678).

Item response theory (IRT): sample-independent evidence

The results of IRT (Table 2) show that the items were monotonic, indicating that variability is linked to a single construct (reading and writing; values equal to or greater than 0.6). The borderline value of item 15 (0.59) was disregarded, as the infit and outfit statistics were good. On residual analysis, items 1 and 10 showed borderline misfit on both infit and outfit measures (item 1: outfit = 0.48, infit = 0.53, $x^2 = 0.04$; item 10: outfit = 1.64, infit = 1.28, $x^2 = 0.04$). Furthermore, the thresholds of item 11 (never-rarely = -0.90; rarely-sometimes = 0.63; sometimes-often/always = -0.12) did not present an adequate relationship between the frequency of symptom manifestations and the respondents' skill as calculated by IRT.

The scale was able to measure the reading and writing ability of children from the 10th percentile upward (Table 3). By overlaying the difficulty parameters (Table 2) on the skill curve (Figure 1, supplementary material), it can be observed (Table 3) that the most informative part of the scale lies between the 10th and 90th percentiles, i.e., between scores 16 and 58.

Convergent validity: relationship between the SRW and reading and writing tests

Spearman's correlation coefficients were moderate (≥ 0.40 to < 0.60) for 11 of the 17 external variables (Table 4).

Table 2. Residuals analysis of the partial credit model using four response categories

Item	Outfit	Infit	Monotonicity	p.S_X2	Never - Rarely	Rarely - Sometimes	Sometimes - Often/Always	Difficulty
Item.1	0.48	0.53	0.78	0.04	0.02	0.66	1.14	0.60
Item.2	0.41	0.57	0.77	0.08	0.15	0.65	0.79	0.52
Item.3	1.23	1.18	0.71	0.56	0.31	0.54	2.16	1.00
Item.4	0.57	0.72	0.75	0.65	-0.07	0.99	2.04	0.98
Item.5	0.67	0.80	0.72	0.32	-0.35	0.89	1.69	0.75
Item.6	0.53	0.75	0.77	0.05	0.85	2.42	2.40	1.89
Item.7	0.51	0.60	0.77	0.37	-0.44	0.63	1.80	0.66
Item.8	1.15	1.16	0.68	0.17	-0.89	0.53	0.88	0.17
Item.9	0.63	0.68	0.77	0.29	-1.58	-0.06	0.89	-0.25
Item.10	1.64	1.28	0.62	0.04	-0.80	-0.56	0.97	-0.13
Item.11	1.18	1.28	0.67	0.26	-0.90	0.63	-0.12	-0.13
Item.12	1.80	1.28	0.60	0.16	-0.01	1.21	1.26	0.82
Item.13	0.66	0.90	0.71	0.02	-0.07	0.61	1.40	0.65
Item.14	0.86	0.98	0.72	0.03	-0.97	0.26	1.95	0.41
Item.15	1.19	1.46	0.59	0.47	-0.76	1.53	2.52	1.10
Item.16	0.77	0.87	0.69	0.11	-0.43	0.88	1.83	0.76

Caption: p.S_X2 = significance. Bold font denotes misfit of outfit and infit values

Test–retest reliability: stability of evaluations made with the SRW over time

Assessment of temporal stability was performed by correlating the latent trait estimates of the first and second data collections ($r_s = 0.80, p < 0.0001$). There were no significant differences

between the mean estimates obtained at the two time points (mean difference = -0.001, $t(29) = -0.025324, p = 0.98$).

DISCUSSION

The results of factor analysis demonstrate that the scale assesses a single construct, i.e., the reading and writing aspects measured in this test behaved as a single skill. Thus, it is evident that different cognitive skills, such as executive function and attention, are not interfering with the findings of the scale.

It is well known that reading and writing involve distinct cognitive processes, each with its own peculiarities⁽²⁷⁾, and can even be learned separately. Therefore, we believed that the SRW would be composed of two skill constructs. However, because these two skills are highly interdependent⁽¹⁵⁾, the items on the scale were unable to measure their distinct features, at least in third-graders.

Analysis of the internal structure of the scale allowed us to obtain an indicator of the reliability of the scale and of the symptoms it investigates⁽¹¹⁾. However, it is worth noting that such high Cronbach's alpha coefficients as found in this analysis can represent item redundancy, i.e., the presence of very similar items on the scale. To investigate this issue, we performed IRT analyses, which indicate the level of skill that a given item evaluates^(11,12) without overlap.

First, the results of IRT confirmed that the SRW evaluates a single construct (monotonicity values equal to or greater than 0.6). Analysis of residuals indicated that items 1 (Takes longer than peers to read *words*) and 10 (Is better at telling a story aloud than at writing it down) showed misfit of outfit and

Table 3. Scoring norms and distribution of the sample

Distribution of scores by percentile		
Percentile	Crude score	IRT-adjusted score
0%	16	-3.747
5%	16	-3.747
10%	16	-3.747
15%	19	-2.069
20%	20.6	-1.585
25%	24	-0.999
30%	26.9	-0.604
35%	27.55	-0.502
40%	31	-0.189
45%	32	-0.098
50%	33.5	0.074
55%	36	0.236
60%	36	0.313
65%	39.9	0.606
70%	44.1	0.937
75%	47	1.118
80%	51	1.431
85%	54.05	1.811
90%	58	2.392
95%	61.7	3.311
100%	64	4.942

Caption: Crude: total SRW score; IRT: adjusted for children's skill level; Percentile: 0% = no difficulty; 100% = maximum difficulty

Table 4. Correlations between SRW results and reading and writing tasks

			School 1	School 2	School 3	School 4	School 5	School 6	Overall
Dictation, Balanced - Errors	PGC		0.44	0.74	0.63	0.33	0.85	0.84	0.63
	SCR		0.59	0.46	0.74	0.24	0.84	0.84	0.70
	CCR		0.50	0.69	0.23	0.34	0.008*	0.58	0.30
	IR		0.01	0.68	0.34	0.35	0.84	0.60	0.44
	Total		0.37	0.71	0.62	0.34	0.86	0.91	0.65
Letter Copying	Total		0.007*	-0.66	-0.29	-0.01*	-0.91	-0.63	-0.44
	Errors		0.43	-0.11*	0.24	0.02*	-0.46	-0.18*	0.015*
	Speed		0.007*	-0.66	-0.29	-0.01*	-0.91	-0.63	-0.41
Reading Comprehension	Spontaneous		0.22	-0.55	-0.35	-0.85	0.03*	-0.72	-0.38
	Guided		0.18	-0.48	-0.19	-0.84	-0.35	-0.68	-0.32
	Time (s) per c, guided		-0.34	0.57	0.31	0.68	0.69	0.84	0.43
Reading Aloud (Words and Pseudowords)	Speed		0.18	-0.73	-0.17	-0.48	0.87	-0.61	-0.29
	Total		0.52	0.53	0.31	0.46	0.79	0.82	0.52
	Words		0.41	0.74	0.38	0.56	0.55	0.85	0.50
	Pseudo		0.55	0.27	0.21	0.16	0.66	0.53	0.34
	Errors/Speed		0.37	0.65	0.40	0.61	0.99	0.88	0.58
Mean (SD)	Speed		0.19	-0.89	-0.38	-0.50	-0.83	-0.82	-0.56

Caption: *Denotes nonsignificant correlations. 1) Error types in the Balanced Dictation task: a) Phoneme-to-grapheme conversion (PGC); b) Simple contextual rules (SCR); Complex contextual rules (CCR); Language irregularity (IR); Total errors in the task (Overall); 2) Letter copying: a) Total letters copied (Total); b) Total spelling errors (Errors); Character speed per minute (Speed); 3) Reading comprehension: a) Number of propositions spontaneously evoked out of 18 (Spontaneous); b) Number of correct answers to questions based on the text, maximum score 10 points (Guided); c) Coefficient between the total time, in seconds, and the number of correct answers to the questions addressed (Time (s) per c, guided); d) Reading speed, words per minute (Speed); 4) Reading Aloud (Words and Pseudowords): a) Total errors in the task (Total); b) Total errors in words (Words); c) Total errors in pseudowords (Pseudowords); d) Total errors in the task as a function of speed (Errors/Speed); e) Reading speed, words per minute (Speed)

infit statistics. However, due to the clinical value of these items and because the misfits were considered borderline, both items were retained.

Analysis of symptom frequency (Never-Rarely; Rarely-Sometimes; Sometimes-Often/Always) revealed a discrepancy between the frequency of symptom manifestation and respondent skill as calculated by the IRT for item 11 (Takes longer than peers when copying (e.g., from the blackboard)). This may indicate that more than one variable interferes with the process of responding to this item⁽²⁸⁾, such as the attention factor. Nevertheless, exclusion of this item was deemed unnecessary, since all other measures resulting from the item showed good values.

One more measure of reliability of the scale was obtained by IRT analysis. By overlaying the difficulty parameters of the items (Table 2) on the information curve (Figure 1, supplementary material), we found that evaluation was most accurate in the “intermediate-high”, “intermediate”, and “intermediate-low” skill ranges. The analysis also showed that there was no overestimation and little underestimation of skill ranges as measured by the SRW.

As shown in Table 3, percentiles equal to or less than 10% and those equal to or greater than 90% were not in the most informative region of the IRT information curve (Figure 1 - supplementary material); therefore, the level of ability of children in these ranges is poorly assessed by the scale. Notably, the scale is unable to detect skill differences up to the 10th percentile, with an IRT value of -3.747. This result suggests that, above a certain degree of skill, all students would be classified in the “Never” category. The analysis also suggests that those students ($n = 9$) who obtain scores equal to or greater than 58 points (90th percentile) should be referred for diagnostic evaluation due to their degree of difficulty reading and writing. These children would be at risk for DD. The percentage of scores in this range (approximately 9% of the sample) corroborates the prevalence of DD reported in different countries (5-10%)⁽²⁹⁾.

Regarding convergent validity, 11 of the 17 variables showed moderate correlation. There is no consensus in the literature as to the appropriate magnitude of correlation for convergent validity⁽³⁰⁾. Urbina⁽¹¹⁾ notes simply that the correlation must be strong, while DeVon et al.⁽³⁰⁾, in their review, indicate that values higher than 0.50 are infrequent, as it is often impossible to find a validated task with the same specificities as the construct of interest to perform the correlation.

Regarding the tests performed, we must make note of a problem with the comparison criteria used for the present study. There is no gold-standard instrument for assessment of reading and writing in Brazil. The most widely used assessment tool, cited in 478 studies (Google Scholar, 2016), is the School Performance Test (*Teste de Desempenho Escolar*, TDE)⁽³¹⁾. However, the version available at the time of the study was constructed more than 20 years ago, and is now outdated⁽³²⁾. None of the tests used in this study have been validated, and only one (the Balanced Dictation task) has had norms described⁽¹⁵⁾.

Based on the arguments advanced by DeVon et al.⁽³⁰⁾, we believe that SRW results have an important association with actual performance on reading and writing tasks. Convergent validity had to be assessed with schools as the unit of analysis,

as there were major differences in the average performance on reading and writing tasks⁽⁷⁾ across institutions. Because of this variation, given a student who made 60 spelling errors on the Balanced Dictation task⁽¹⁵⁾, teachers from different schools would probably score this same student differently on the corresponding SRW item.

The differences found in average reading and writing performance scores in this research may be largely related to differences in methodology and syllabus across the sampled schools. The Brazilian National Curriculum Guidelines for Basic Education⁽³³⁾ are limited to methodological principles (interdisciplinary and problem-based learning), without specifying what they are or how they should be worked on. It is thus up to each teacher to choose the best teaching methodology for their group; therefore, strategies for presenting content to the class may vary from educator to educator, thus leading to differences in student performance.

The SRW investigates reading and writing, skills that involve different cognitive processes but are interdependent, and can thus be compared to a battery of tests. For instance, all comprehension tasks are essentially related only to item 7 of the scale, whereas those related to the Balanced Dictation task have a more intrinsic relationship with item 8. Although factor analysis indicates that reading performance and writing performance correlate strongly with one another, to the extent of being considered representative of a single factor, the greater specificity of individual tasks may have decreased the correlation strength, as observed when comparing test batteries to an isolated task⁽¹¹⁾.

Only one variable was uncorrelated with the scale: the number of errors made when copying. We presume this occurred because children with persistent difficulties, being aware of their problem, create strategies to avoid mistakes. Thus, there is no impact on accuracy, but rather on the speed with which they complete the task.

The weakest correlation was that of reading speed. As there are no parameters for assessment of reading fluency in school settings⁽³⁴⁾, this evaluation is entirely subjective. The low correlation with comprehension scores can be a reflection of conceptual flaws about comprehension and of the screener instrument. In this line, studies have shown gaps in teacher knowledge regarding the processes that underlie the development of reading⁽³⁵⁾.

The subjective nature of teachers' perceptions of differences between students may also be associated with the strength of the correlations found in this study. The strength of correlation varied widely across institutions, even ranging from positive to negative. Teacher training and seniority may be directly involved in this difference between institutions, as well as other social and demographic variables of schools.

Finally, regarding the temporal stability (test-retest reliability) coefficient, optimal values are generally defined as those equal to or greater than 0.90⁽¹¹⁾; however, values as low as 0.80 are considered acceptable⁽¹²⁾. Several factors may explain subpar values.

Issues such as the time elapsed between assessments and a potential decrease in participant motivation when retaking a test

interfere with this correlation⁽¹²⁾. An interval of 15 to 30 days between measurements is recommended. However, returns for the retest step were only received 2 to 4 months after the initial evaluation; reasons included a delay in delivery, as initially only one teacher had agreed to participate, which also appears to indicate a reduction in motivation among the sample of teachers.

CONCLUSION

The processes described in this article provide evidence of the validity of the SRW screening instrument (Appendix A), according to the principles set forth by the American Psychological Association and the *Conselho Federal de Psicologia*⁽³⁶⁾. Although the SRW was developed to assess the reading and writing skills of students from the first to the fifth grade, assessment of its validity was restricted to third-graders.

As noted in the introduction, this was a deliberate choice, considering the diagnostic criteria for DD and the provisions of the *Pacto Nacional pela Alfabetização na Idade Certa* (PNAIC, National Pact for Literacy at the Right Age), at the time of assessment, and of the current BNCC. However, in 2019, conceptual changes led to a major update of the Brazilian National Literacy Policy⁽³⁷⁾. The new policy aims to ensure that children are able to read and write simple texts by completion of the second grade of elementary school. This new concept in no way invalidates the present study or the SRW. Its items continue to represent the symptoms of DD, and this reconceptualization does not require any changes to the statistical analyses, especially those referring to the SRW items and their results, which were shown to correlate with performance on reading and writing tasks. In addition, as previously noted, the BNCC continues to regard the third grade as a literacy milestone, and it is in the third grade that the national literacy assessment is carried out.

By demonstrating that the proposed screener actually measures what is sets out to measure, this study fills an important gap in the field. The SRW provides physicians, speech therapists, psychologists, and educators with a tool which yields reliable evidence for the identification of students at risk of DD. The SRW can also be used in research settings, by investigators who wish to select third-graders with and/or without impairments in the development of reading and writing skills for study samples. Finally, the scale can serve as a population-level screening instrument for research purposes.

ACKNOWLEDGMENTS

The authors would like to thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior; process 23038.002530/2013-93) for financing this project. We would also like to thank the referees who evaluated the scale and Euclides José Mendonça Filho for contributing with the statistical analyses.

REFERENCES

1. APA: American Psychiatric Association. DSM-5: manual diagnóstico e estatístico de transtornos mentais. Porto Alegre: Artmed; 2014.
2. IBGE: Instituto Brasileiro de Geografia e Estatística. Taxa de analfabetismo das pessoas de 15 anos ou mais de idade, por sexo - Brasil - 2007/2014 [Internet]. Rio de Janeiro: IBGE; 2016 [citado em 2016 Jul 27]. Disponível em: <http://brasilemsintese.ibge.gov.br/educacao/taxa-de-analfabetismo-das-pessoas-de-15-anos-ou-mais.html>
3. Costa AC, Toazza R, Bassoa A, Portugez MW, Buchweitz A. Ambulatório de Aprendizagem do Projeto ACERTA (Avaliação de Crianças Em Risco de Transtorno de Aprendizagem): métodos e resultados em dois anos. In: Salles J, Haase VG, Malloy-Diniz LF, editores. Neuropsicologia do desenvolvimento: infância e adolescência. Porto Alegre: Artmed; 2015. p. 151-8.
4. Colenbrander D, Ricketts J, Breadmore HL. Early identification of dyslexia: understanding the issues. *Lang Speech Hear Serv Sch*. 2018;49(4):817-28. http://dx.doi.org/10.1044/2018_LSHSS-DYSLC-18-0007. PMID:30458543.
5. Torres D, Ciasca S. Correlação entre a queixa do professor e a avaliação psicológica em crianças de primeira série com dificuldades de aprendizagem. *Rev Psicopedag*. 2007;24(73):18-29.
6. Bussing R, Fernandez M, Harwood M, Wei Hou, Garvan CW, Eyberg SM, et al. Parent and teacher SNAP-IV ratings of attention deficit hyperactivity disorder symptoms: psychometric properties and normative ratings from a school district sample. *Assessment*. 2008;15(3):317-28. <http://dx.doi.org/10.1177/1073191107313888>. PMID:18310593.
7. Bassôa A. Construção e evidências de fidedignidade e validade de uma Escala de Leitura e Escrita (ELE) para o rastreio de crianças com dificuldades escolares [dissertação]. Porto Alegre: Pontifícia Universidade Católica do Rio Grande do Sul; 2016.
8. Torgesen JK. The prevention of reading difficulties. *J Sch Psychol*. 2002;40(1):7-26. [http://dx.doi.org/10.1016/S0022-4405\(01\)00092-9](http://dx.doi.org/10.1016/S0022-4405(01)00092-9).
9. Torgesen JK. Individual differences in response to early interventions in reading: the lingering problem of treatment resisters. *Learn Disabil Res Pract*. 2000;15(1):55-64. http://dx.doi.org/10.1207/SLDRP1501_6.
10. Gresham FM, Elliott SN. Inventário habilidades sociais, problemas comportamento e competência acadêmica para crianças - SSRS. São Paulo: Casa do Psicólogo; 2016.
11. Urbina S. Essentials of psychological testing. 2nd ed. New Jersey: Wiley; 2014.
12. Pasquali L. Psicometria: teoria dos testes na psicologia e na educação. 5. ed. Rio de Janeiro: Vozes; 2013.
13. Salles JF. Habilidades e dificuldades de leitura e escrita em crianças de 2ª série: abordagem neuropsicológica cognitiva [dissertação]. Porto Alegre: Universidade do Estado do Rio Grande do Sul; 2005.
14. Saraiva RA, Moojen SMP, Munarski R. Avaliação da compreensão leitora de textos expositivos para fonoaudiólogos e psicopedagogos. São Paulo: Casa do Psicólogo; 2005.
15. Moojen SMP. A escrita ortográfica na escola e na clínica: teoria, avaliação e tratamento. São Paulo: Casa do Psicólogo; 2011.
16. Brown TA. Confirmatory factor analysis for applied research. 2nd ed. USA: Guilford Publications; 2015.
17. Holgado-Tello FP, Chacón-MoscOSO S, Barbero-García I, Vila-Abad E. Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Qual Quant*. 2009;44(1):153-66. <http://dx.doi.org/10.1007/s11135-008-9190-y>.
18. Hair JF, Black WC, Anderson R, Babin B. Multivariate data analysis: a global perspective. 7th ed. Upper Saddle River: Pearson Education; 2010.
19. Masters GN. A Rasch model for partial credit scoring. *Psychometrika*. 1982;47(2):149-74. <http://dx.doi.org/10.1007/BF02296272>.
20. Wright BD, Linacre JM. Reasonable mean-square fit values. *Rasch Meas Trans*. 1994;8:370.
21. Muthén LK, Muthén B. Mplus user's guide. 6th ed. Los Angeles, CA: Muthén & Muthén; 2010.
22. R Core Team. R: a language and environment for statistical computing [Internet]. Vienna: Foundation for Statistical Computing; 2015 [citado em 2016 Ago 27]. Disponível em: <https://cran.r-project.org>
23. Revelle W. Psych: procedures for personality and psychological research [Internet]. Evanston, Illinois: Northwestern University; 2015 [citado em 2016 Ago 27]. Disponível em: <http://cran.r-project.org/package=psych>

24. Chalmers RP. mirt: a multidimensional item response theory package for the R environment. *J Stat Softw.* 2012;48(6):1-29. <http://dx.doi.org/10.18637/jss.v048.i06>.
25. Cattell RB. The scree test for the number of factors. *Multivariate Behav Res.* 1966;1(2):245-76. http://dx.doi.org/10.1207/s15327906mbr0102_10. PMID:26828106.
26. Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika.* 1965;30(2):179-85. <http://dx.doi.org/10.1007/BF02289447>. PMID:14306381.
27. Graham S, Hebert M. Writing to read: A meta-analysis of the impact of writing and writing instruction on reading. *Harv Educ Rev.* 2011;81(4):710-44. <http://dx.doi.org/10.17763/haer.81.4.t2k0m13756113566>.
28. DeMars C. Item response theory. USA: Oxford University Press; 2010. <http://dx.doi.org/10.1093/acprof:oso/9780195377033.001.0001>.
29. Law JM, Vandermosten M, Ghesquiere P, Wouters J. The relationship of phonological ability, speech perception, and auditory perception in adults with dyslexia. *Front Hum Neurosci.* 2014;8(482):482. <http://dx.doi.org/10.3389/fnhum.2014.00482>. PMID:25071512.
30. DeVon HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ, et al. A psychometric toolbox for testing validity and reliability. *J Nurs Scholarsh.* 2007;39(2):155-64. <http://dx.doi.org/10.1111/j.1547-5069.2007.00161.x>. PMID:17535316.
31. Stein LM. TDE: teste de desempenho escolar: manual para aplicação e interpretação. São Paulo: Casa do Psicólogo; 1994. 17 p.
32. Athayde MDL, Mendonça EJD Fo, Fonseca RP, Stein LM, Giacomoni CH. Desenvolvimento do subteste de leitura do Teste de Desempenho Escolar II. *Psico-USF.* 2019;24(2):245-57. <http://dx.doi.org/10.1590/1413-82712019240203>.
33. Brasil. Ministério da Educação. Diretrizes Curriculares Nacionais da Educação Básica. Brasília: MEC, SEB, DICEI; 2013.
34. Komeno EM, Ávila CRBD, Cintra IDP, Schoen TH. Velocidade de leitura e desempenho escolar na última série do ensino fundamental. *Estud Psicol.* 2015;32(3):437-47. <http://dx.doi.org/10.1590/0103-166X2015000300009>.
35. Oliveira FJD, Silveira MIM. A compreensão leitora e o processo inferencial em turmas do nono ano do ensino fundamental. *Revista da FAEEBA.* 2014;23(41):91-104. <http://dx.doi.org/10.21879/faeaba2358-0194.v23.n41.826>.
36. Conselho Federal de Psicologia. Avaliação psicológica: diretrizes na regulamentação da profissão. 1. ed. Brasília: CFP; 2010.
37. Brasil. Ministério da Educação. Secretaria de Alfabetização. Política Nacional de Alfabetização (PNA). Brasília: MEC, SEALF; 2019.

Author contributions

AB(1) participated in all stages of the research: building the protocol, data collection, statistical analyses and manuscript writing; AB (2) supervised the research and contributed to analyses and manuscript writing; ACC and RT contributed to designing the screener and revising the manuscript.

APPENDIX A. SCREENER

Name of student:

SCREENER FOR READING AND WRITING - SRW

TEACHER VERSION

	Never	Rarely	Sometimes	Often/Always
1. Takes longer than peers to read words;				
2. Takes longer than peers to read texts;				
3. Switches letters when reading syllables or words aloud;				
4. Sounds out words before reading them aloud;				
5. Stutters, shakes, blushes and/or repeats words when reading aloud;				
6. "Makes up" words, swaps out words for similar ones, or appears to guess words when reading aloud;				
7. Does not understand what has been read (e.g., does not understand what to do after reading instructions, or does not understand what happened to the characters of a story);				
8. Switches/swaps, omits, or adds letters when writing;				
9. Compared to peers, writes texts that are overly simple, unimaginative, and lacking in detail;				
10. Is better at telling a story aloud than at writing it down;				
11. Takes longer than peers when copying (e.g., from the blackboard ⁽¹⁾);				
12. Has poorly legible handwriting;				
13. Avoids situations that involve reading and writing;				
14. Has trouble rhyming or identifying words that rhyme;				
15. During conversations, often takes a long time to recall the names of familiar people, objects, feelings, or learned content ("it's on the tip of my tongue");				
16. Has trouble memorizing lists or sequences of information (e.g., times tables, months of the year, days of the week).				

The following list describes some specific difficulties that may be displayed by children or adolescents with impaired reading and/or writing. Read each statement and check the box that best describes this student's performance (i.e., how often he/she has exhibited each of the difficulties) in the past 6 months. It is very important that you read the whole screener before answering each item.

1 Or chalkboard.

Please answer the following questions, considering the last 6 months:

1. Compared to his/her peers, is this student's language performance well short of expectations?

Yes No

2. Has this student received been receiving any extra support, or have any adaptations been made to accommodate this student?

Yes No

3. Has this student been making substantial progress in his or her reading and writing skills (e.g., able to read and/or understand longer texts; significant reduction in spelling mistakes; etc.)?

Yes No

4. Use the following field to make any notes you may find relevant: _____

SUPPLEMENTARY MATERIAL

Supplementary material accompanies this paper.

Supplementary Table 01: Factor Loadings for the ELE

Supplementary Table 02. Internal Consistency: Values for the Alpha Coefficient

Supplementary Figure 1: Distribution of ELE scores according to the level of skill calculated using IRT

This material is available as part of the online article from <http://www.scielo.br/codas>