

Revised version of the Scale of Evaluation of Reading Competence by the Teacher: final validation and standardization*

Versión revisada de la Escala de Evaluación de la Competencia de Lectura por el Profesor: validación final y estandarización

Received: June 16, 2015 | Accepted: October 10, 2016

Douglas de Araújo Vilhena*

Universidade Federal de Minas Gerais, Brasil

Ângela Maria Vieira Pinheiro**

Universidade Federal de Minas Gerais, Brasil

ABSTRACT

The original version of the EACOL, a tool for teachers to assess silent and aloud reading of Brazilian 2nd-to-5th-graders, was revised and the resulting instrument was validated and normalized. Method: 72 teachers were asked to answer the revised EACOL and a behavioral questionnaire; 452 pupils performed a test battery composed by seven reading tasks and one general cognitive ability measure. Results: The revised EACOL presented high reliability and moderate-to-strong correlations with all reading variables; cluster analysis suggested three proficiency groups (poor/not-so-good/good readers). Conclusion: in agreement with previous studies, teachers, when provided with sound criteria, can come to reliable evaluations of their students' reading ability. Thus, an improved instrument, with evidence of reliability as well as content, internal and external validity, is offered to allow an indirect assessment of the reading ability of schoolchildren. This instrument can easily be adapted to other Portuguese-speaking countries.

Keywords

reading skills, reading assessment, child assessment, Portuguese language, teacher scale.

RESUMEN

La versión original de EACOL es una herramienta para que los profesores evalúen la lectura silenciosa y en voz alta de los estudiantes brasileños del segundo al quinto año escolar, esta fue revisada, validada y estandarizada. Método: 72 profesores respondieron la escala EACOL y un cuestionario de comportamiento; 452 estudiantes respondieron siete medidas de lectura y una de capacidad cognitiva general. Resultados: la revisión de EACOL mostró una alta confiabilidad y correlaciones de moderadas a fuertes con todas las variables de lectura. Análisis de clusters sugirió tres grupos de competencia (lector de baja/media/ alta). Conclusión: de acuerdo con estudios anteriores, los profesores pueden hacer evaluaciones confiables de la capacidad de lectura de sus estudiantes, cuando se proporciona criterios operacionales. De esta manera, se ofrece un instrumento mejorado para evaluar indirectamente la lectura de niños, con evidencias de fiabilidad interna y externa validez de contenido. Este instrumento se puede adaptar fácilmente a otros países de lengua portuguesa.

Palabras clave

*Doctoral Student in Psychology. E-mail: douglasvilhena@gmail.com

**Full Professor. Ph.D. in Cognitive Psychology, University of Dundee, Scotland. E-mail: pinheiroamva@gmail.com

habilidades para la lectura, evaluación de lectura, evaluación de niños, lengua portuguesa, escala de profesores.

How to cite: Vilhena, D.A., & Pinheiro, A.M.V. (2016). Revised version of the Scale of Evaluation of Reading Competence by the Teacher: final validation and standardization. *Universitas Psychologica*, 15 (4). <http://dx.doi.org/10.11144/Javeriana.upsy15-4.efvs>

Introduction

According to the Literacy Initiative for Empowerment (LIFE; United Nations Educational, Scientific and Cultural Organization [UNESCO], 2007), education is a human right and a public good that enables access to information about health, the environment, the world of work and, most importantly, how to learn throughout life. This assertion is of particular relevance in the Brazilian context as only 56.1% of children are fully literate at 8 years of age (Todos pela Educação, 2013) and 11% of young people aged 15–24 remain functionally illiterate (Instituto Paulo Montenegro, 2011).

Given this situation, a proactive approach is needed. Nothing justifies waiting for students to fail, as the focus of literacy education should be on the prevention of reading problems rather than on remedial intervention. Early screening for reading difficulties can be appropriately done by elementary school teachers, who are undeniably one of the most important sources of information about their students. Snowling, Duff, Petrou, Schiffeldrin, and Bailey (2011) asserted that teachers evaluations of their students reading skills, when criterion-referenced assessments are made available, can be as good as those of most formal tests. It is possible that with clear criterion, the teachers' judgments are less influenced by factors beyond the school performance itself, such as gender, social behavior and socioeconomic characteristics (Bennett, Gottesman, Rock, & Cerullo, 1993; Soares, Fernandes, Ferraz, & Riani, 2010).

In Brazil, there is a lack of instruments with validity and precision to guide teachers in an initial categorization of the reading abilities of their students. The development of the Scale of Evaluation of Reading Competence by the Teacher (in Portuguese, Escala de Avaliação da Competência em Leitura pelo Professor, or EACOL) (Pinheiro & Costa, 2015) is an initiative to fill this gap. However, previous studies identified issues indicating that the scale needed revision (Lúcio & Pinheiro, 2013). In this paper, we present the improvements in EACOL in response to these issues, followed by validation and standardization of the resulting final version of the scale.

The EACOL

Pinheiro e Costa (2015) provided evidence of content validity to EACOL by the judgment of specialists of a set of descriptors of good, not-so-good and poor Reading Aloud and Silent Reading behaviors that could be recognized by the teacher. Reading Aloud items measure speed and accuracy in word recognition, prosody, and comprehension; whereas Silent Reading items measure comprehension and the capacity for synthesis. After this procedure, two scales were created: a) Form A, with 23 items for 2nd-graders (in elementary school), who are at or near the beginning of the literacy process, with an average age of 7 years; and b) Form B: with 27 items for students from 3rd to 5th grade, at the later stage of literacy learning and also for readers already literate, with an approximate age of 8–11 years. The study of Pinheiro e Costa remained only in the theoretical validation bases, as there was no direct assessment of the students.

A first internal and external validation of the EACOL's Form B was carried out by Cogo-Moreira, Ploubidis, Brandão de Ávila, Mari, & Pinheiro (2012). Using the statistical Latent Class Analysis method, the three types of readers expected by the authors of the EACOL (good, not-so-good, and poor readers) were found. Out of 27 items of the Form B, only two items showed an overlap – *Reads too slowly*

or too quickly and *Reads words correctly* – suggesting that they required revision. The study established a concurrent validation with word naming tasks only, as text comprehension was not evaluated. On the other hand, psychiatric behaviors and non-verbal intelligence measures provided evidence of discriminant validity.

In spite of the general good quality of the instrument evidenced in Cogo-Moreira et al. (2012) study, there were two points of concern about it. The first refers to the number of items actually filled by the teachers, and the second to the awareness that the instrument could be more attractive to the teachers if it were to be shortened. Taking the first point, later scrutiny of the data revealed that a significant number of items had not been answered. It was reasoned, then, that such a result could have been due to the dichotomous nominal level of response demanded by the instrument: “Yes” and “No”, as in this case a teacher may be prone to waive an answer if he or she is not pleased with either alternative. Another problem with binary choice is that the respondents tend to favor the positive alternatives rather than the negative ones (Emmerich, Enright, Rock, & Tucker, 1991). Thus, in an attempt to obtain more control over the answers given by teachers and to avoid the problems associated with binary options, the alternative “I do not know” was added as a third option.

As for the second concern, in order to make the instrument shorter, it was realized that the set of items describing the not-so-good reader category [e.g., *Sometimes makes mistakes when (...)*, *Not always is able to identify (...)*, and *Presents some difficulty in (...)*] could be excluded and that the idea of a behavior that sometimes occurs and sometimes does not would be replaced by the option “sometimes”, which would be included within the response alternative of the scale. In this way, only the items requiring a “yes” or “no” response that respectively describes the good and the poor reader would be kept, which required a further change not only in the structure of the scale, but also in its scoring criterion.

Finally, again inspired by studies evaluating the reliability of multiple-choice answers

(e.g., Verbic, 2012), we replaced the options “Yes” and “No” with “True” and “False” to avoid misinterpretation of items with negative statements. For example, on the item *Not always able to identify the subject from the title and vice versa*, while a “Yes” answer indicates a poor reader, a “No” answer indicates a good reader. In such cases, the teacher may erroneously assign a “Yes” to a good performance or a “No” to a poor performance, which would lead to an inaccurate judgment of the child’s ability.

To summarize, in this revision, EACOL underwent the following modifications: a) replacement of “Yes” by “True” and “No” by “False”; b) replacement of the binary option for answers by four choices: “True”, “False”, “Sometimes”, and “I do not know”; c) exclusion of the set items about the not-so-good reader due to the new response format; d) addition and revision of other items; and e) identification and selection of the best scoring criterion to the new format of the scale. These modifications were tested and evidence of validity and of reliability were provided, as well as standardization of the resulting revised version, being this the first validation study for the Form A.

Method

Participants

To evaluate whether the teacher’s judgment is as reliable as a direct reading assessment, the cognitive functions of 2nd-to-5th-graders were evaluated to provide evidence of concurrent validity (see Table 1 for the pupils’ sociodemographic distribution). The sample (452 students and 72 teachers across 8 state schools) was tested from November to December 2013. Only six students were randomly selected in each classroom. The institutions were arbitrarily chosen from a document provided by the State Secretary of Education, stratified over the districts in Belo Horizonte city.

Schools, teachers, pupils and their guardians signed an informed consent form for the

research. The assessments were administered during school hours, in a quiet room in the institution. All participants provided informed consent, and the Ethical Committee from the Universidade Federal de Minas Gerais approved the study (Certificate of Appreciation Presentation to Ethics [Certificado de Apresentação para Apreciação Ética; CAAE]: 17754514.6.0000.5149).

Table 1
Items and scores of the EACOL – Scale of Evaluation of Reading Competence by the Teacher

Age in years	Sex		Grade				Total
	Male	Female	2 nd	3 rd	4 th	5 th	
7	24	15	39				39
8	48	61	57	51	1		109
9	61	48		41	65	3	109
10	62	76		3	57	78	138
11	26	28		1		53	54
12	1	0				1	1
13	2	0				2	2
Total	224	229	96	96	123	137	452

Form A (2nd grade) contains only the underlined sentences and Form B (3rd-5th grade), both sets: underlined and non-underlined sentences. Each item is followed by the alternatives: True, False, Sometimes, I do not know.
Source: own work

Instruments

The revised version of EACOL is composed of two forms (A and B) that differ from its original version in their number of items and in its content. Form A consists of 15 items and Form B of 21 items (against 23 and 27 items, respectively, in the original version of the instrument). In front of all items are the alternative answers “True”, “False”, “Sometimes”, and “I do not know”.

Child behavior was assessed by the Strengths and Difficulties Questionnaires (SDQ), which is a brief behavioral screening questionnaire for 4–16-year-olds (Goodman, 1997; Cury & Golfeto, 2003; Saur & Loureiro, 2012). This study used the single-sided Brazilian version, with scoring for teachers (Goodman, 2005), composed by 25 items divided into 5 scales: emotional symptoms (anxiety/mood), conduct problems (aggression/delinquency), hyperactivity/inattention, peer relationship problems (withdrawn/social problems), and prosocial behavior (empathy/positive relations).

The Word Reading Task (WRT) and the Pseudoword Reading Task (PWRT) are Reading Aloud instruments each consisting of 88 words and 88 pseudowords printed on an A4 page, font Ariel size 14 (Pinheiro, 2013). The psycholinguistic variables for the words were a) frequency of occurrence (high vs. low), b) bidirectional regularity (regular and irregular words according to grapheme-to-phoneme correspondence and vice versa), and c) length (short, medium, and long words). The pseudowords were constructed with the same orthographic structures and stimulus length used in the word task.

The Reading Test – Sentence Comprehension (Teste de Leitura – Compreensão de Sentenças, TELCS) was used to evaluate the silent reading efficiency (Vilhena, Sucena, Castro, & Pinheiro, 2016). It consists of 36 incomplete and isolated sentences, each followed by five words as alternative fill-in-the-blank answers. The child’s task is to select, in up to 5 minutes, the best word to give meaning to each sentence.

Another instrument used to evaluate the silent reading was the Text Reading Comprehension subtest (PROLEC-text), which is part of the PROLEC (Provas de Avaliação dos Processos de Leitura [Reading Processes Assessment Battery]; for Capellini, Oliveira & Cuetos, 2012). It consists of four short texts to investigate students’ ability to answer sixteen literal questions.

General cognitive ability was measured using Raven’s Coloured Progressive Matrices Test (CPM) (Angelini, Alves, Custódio, Duarte, &

Duarte, 1999). It evaluates analogic reasoning, or the ability to infer relations between objects or elements (Pasquali, Wechsler, & Bensusan, 2002). It is used mainly for children between 5 and 11 years, and consists of 36 items divided into three sets of 12 (A, Ab, B) arranged in inter- and intraset according to increasing difficulty. The task is to select the best option to, fill-in the gap, among six alternatives printed beneath.

Procedures

Each teacher was asked to answer, during a period of one week, the EACOL and SDQ for six students only. All instruments answered by students were administrated on the same day, in two sessions, each lasting on average 15 minutes. Whereas in the first session, groups of up to 10 children were collectively submitted to both TELCS and CPM, in the second, each individual child was presented with the pair WRT and PWRT (in random order), followed by the PROLEC-Text.

To guarantee EACOL's internal consistency, two exclusion criteria were established to control possible incongruence and/or unjudgeability on a given scale: a) opposing items answered more than twice, and b) presence of four or more items not answered or "I do not know" responses. Either of these criteria would lead to the exclusion of that scale from the sample.

The WRT and the PWRT tests were administrated in sequence, but in a random order. Participants were asked to read aloud each item of each test card, starting from the first row from to right. The reading time and errors were registered by the applicator. On both instruments, two measures were used: accuracy, which is the total number of correctly read words or pseudowords, and accuracy rate, which is the total number of correct words or pseudowords read per minute.

The TELCS was administered with a training phase composed of four items, with the first two answered collectively after being read aloud by the researcher and the other two individually, via silent reading. The remaining 36 items were also

read in silence by each child, however, as quick as possible within a maximum of five minutes and with no assistance granted. The scoring of the test consisted of one point for each correct answer and zero for the incorrect or omitted ones.

The PROLEC-Text's stories were administrated in a fixed order, after the following statement: "I will display a small text for you to read. Read it carefully because after you finish I will ask you some questions about them". The participant was asked to read each story, in silence, without time limit, and to respond orally to open questions (also made orally), immediately after reading each text. No rereading was allowed.

The CPM was individually administrated to 2nd year students and the collective form was used for students from grades 3 to 5. It was presented as a puzzle game: the first two items were introduced collectively and explicitly, with subsequent items answered without assistance. There was no time limit. No child spent more than 12 minutes to complete the test.

Statistical analyses

All analyses were performed using IBM SPSS Statistics version 21.0. Due to the diversity in EACOL's item structures, all data were transformed to represent only a Likert-type scale from negative to positive. A hypothetical-deductive method using a Pearson bivariate correlation with all the instruments was applied to determine which was the best scoring criterion for the alternatives of each item from the EACOL. Four scoring hypotheses were tested: a) bad reading: 0, not-so-good: 1, good: 2; b) bad reading: 0, not-so-good: 2, good: 3; c) bad reading: 0, not-so-good: 1, good: 3; d) bad reading: 0, not-so-good: 0, good: 2. The answer "I do not know" was assigned the same score as those corresponding to the category "not-so-good-readers". Cronbach's alphas were calculated to estimate the reliability of EACOL's Forms A and B. A hypothetical-deductive method can confirm if the removal of any item can alter the alpha and the concurrent validity correlations.

As EACOL evaluates reading competence as a whole, dimension reduction by principal component analysis (Carreira-Perpiñán, 1997) was used to incorporate all four reading instruments into a robust reading measure, from here on called the General Reading Composite. A reliability analysis indicated the use of the raw scores from the PROLEC-text, TELCS, Word Reading Task accuracy rate, and Pseudoword Reading Task accuracy. This integration of measures enables us to represent the child's reading performance with a single variable.

A two-step cluster analysis was used to verify the number of mutually exclusive latent groups in the sample. The only variables used were the score for each item in EACOL. This method is a scalable cluster analysis algorithm designed to handle large data sets in two steps: 1) pre-cluster the cases into many small sub-clusters; 2) cluster these sub-clusters into the desired number of clusters. The log likelihood distance measure was used, with subjects assigned to the cluster leading to the largest likelihood. The Bayesian information criterion (BIC) was established to compare the number of latent classes, a comparison in which small values correspond to better fit. Differences in the sample were compared according to cluster membership using a univariate Analysis of Variance (ANOVA) test. For all tests performed, the significance level was set at 0.05.

Results

Item revision

Due to the addition of the alternative "Sometimes", the following eight items, descriptors of the not-so-good-readers, were removed in both Form A and Form B: a) *Sometimes reads and cannot retell what was read* ; b) *Reads too slowly or too quickly* ; c) *Sometimes makes mistakes when reading "new" words* ; d) *Sets the tone of interrogation and/or exclamation only in the word that precedes the punctuation mark* ; e) *Slows the rhythm of reading when "new" words are encountered,*

needing to spell them out ; f) *Not always able to identify the subject from the title and vice versa* ; g) *Does identify characters and places, but has some difficulty identifying main ideas without a second reading* ; and h) *Has some difficulty in orally summarizing what was read.*

Within these, the item (b) *Reads too slowly or too quickly*, was one of the two that showed poor discrimination in Cogo-Moreira et al. (2012). The other, *Reads words correctly*, was also removed for being rather vague. Finally, the last excluded item was a descriptor of a poor reader (*Says "I do not know" when encounters a new word*), since there is another item in the scale that deals with reading of new words and to avoid confusion with the new alternative answer "I do not know".

In contrast to these 10 removed items, 5 others were added (one in Form A and the remainder in Form B). This was thought to be necessary to increase the number of descriptors of the ability of the readers and to maintain the power of the scale. The descriptor of poor reading *Reads with difficulty "known" words* was added to Form A. The following items were added to Form B: a) *Reads clearly, without "stumbling" or "swallowing" syllables. Someone who hears can understand what is being read* ; b) *Has great difficulty in Reading Aloud* ; c) *Reads without pronouncing words or without moving the lips, only moving the eyes* ; and d) *Cannot read without movements of the lips or without pronouncing the words.*

Finally, the item *Reads "new" and invented words quickly* was changed into *Reads "new" words correctly*. The omission of "invented words" was motivated by the fact that pseudowords are rarely presented to students in school. Equally, the alteration of *quickly* into *correctly*, was motivated by the expectation that although automatized reading of both known and new words is the end point in literacy learning, correct word reading, especially of new words, is achieved before the gain of speed.

The original version of EACOL was structurally divided into Reading Aloud and Silent Reading subscales, but this separation did not show to be justifiable in the current version

due to the reduction of items (although the items were statistically analyzed individually). In addition, in the Reading Aloud subscale there are two items that evaluated reading comprehension (e.g., Seems to have understood what was read when asked about the text read), and in the Silent Reading subscale there are eight items that expressed behaviors that are not specific to the condition of silent reading. Rather, these behaviors can be assessed in either reading aloud or silent condition (e.g., Does not identify characters, places, or main ideas; Is able to choose a title for passages with no title or even give an alternate title for titled passages).

Validation

On the selection of the scores for EACOL, the strongest correlations were with the first hypothesis (the first scoring criterion). This was the hypothesis under which predictors of poor readers score zero, predictors of good readers score two points, and both predictors of not-so-good-readers (alternative “Sometimes”) and “I do not know” score one point (see Appendix).

In Form A, the Cronbach’s Alpha suggested that the removal of item 5 (*Does not identify characters, places, or main ideas*) would increase the alpha by 0.004. This suggestion was confirmed by the consistent weak correlations of item 5 ($r = 0.244$) with all reading measures. Finally, the total score (sum of both subscales minus the aforementioned item 5) has an alpha of 0.935, demonstrating the strong internal consistency reliability of EACOL’s Form A. In the further analysis of Form A, item 5 will not be considered. The same internal validity test was performed on Form B, that demonstrated a strong Cronbach’s alpha ($\alpha = 0.958$), with a loss in alpha with the removal of any item.

For concurrent validity, to attest to what extent the evaluations of teachers agree with the actual performance of children, correlations were calculated between the scores of EACOL and all reading measures (see Table 2). Forms A and B had correlation ranges with the reading measures of 0.544–0.737 and 0.484–0.688, respectively.

Moderate correlations were found with the General Reading Composite ($r = 0.737$ and 0.688). Unlike in Cogo-Moreira et al. (2012), Form B was significantly correlated ($p < 0.0001$) with CPM ($r = 0.37$) and with the total score of the SDQ ($r = -0.48$). Form A also demonstrated weak correlations ($p < 0.0001$) with CPM ($r = 0.26$) and with all SDQ negative behaviors subscales.

Table 2
Pearson correlation between EACOL, reading, general cognitive ability, and behavior

Measures	Form A	Form B
WRT (accuracy measure)	0.670**	0.484**
(accuracy rate)	0.662**	0.587**
PWRT (accuracy measure)	0.631**	0.522**
(accuracy rate)	0.647**	0.548**
<u>TELCS</u>	0.650**	0.649**
<u>PROLEC-text</u>	0.544**	0.505**
General reading composite	0.737**	0.688**
CPM	0.263**	0.370**
Prosocial behavior	0.158	0.278**
Emotional symptoms	-0.328**	-0.244**
Conduct problems	-0.275**	-0.401**
Hyperactivity/inattention	-0.505**	-0.490**
Peer relationship problems	-0.258*	-0.331**
Total negative behaviors	-0.477**	-0.483**

TELCS: Reading Test – Sentence Comprehension;
WRT: Word Reading Task; PWRT: Pseudoword Reading Task; PROLEC-text: PROLEC Text Comprehension subtest;
CPM: Coloured Progressive Matrices scores.

The four underlined variables combined form the General Reading Composite.

Note. * $p < 0.05$ (2-tailed), ** $p < 0.01$ (2-tailed).

Source: own work

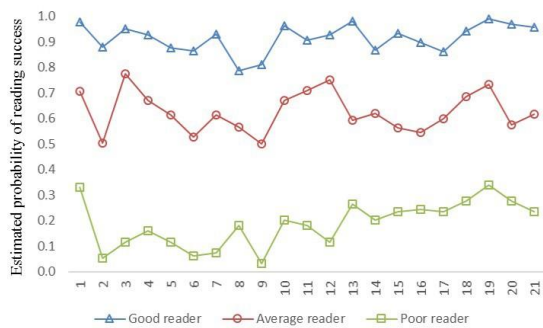
As expected, the two-step cluster analysis suggested a good fit-model with the following three classes for Form B: poor ($n = 47$), not-so-good ($n = 119$), and good readers ($n = 184$). As seen in Figure 1, a clear three-class group

structure is therefore supported, considering both empirical and theoretical elements, with an estimated probability axis scale from 0 (reading disability) to 2 (good reading ability), with no item overlapped. An univariate Analysis of Variance confirmed that all three groups presented significant distinctions from one another on EACOL Total Scores, $F(2, 347) = 1312.7$, $MSE = 14.4$, $p < 0.00001$. The cluster analysis for EACOL's Form A demonstrated the

same pattern as that for Form B, with no item overlap.

Figure 1

Two-step cluster analysis for all 21 items from EACOL's Form B.



Descriptive analysis

No answered scale was eliminated due to internal inconsistency (opposing items answered more than twice) or incapability/difficulty of judgment by the teacher (four or more items answered as “I do not know”). Although the alternative “I do not know” was chosen in just 1% of the possible cases, in 12% of the questionnaires there was at least one answer for this category. Another 1% of the scales returned with at least 1 item without answer; these items were scored with the same value as “I do not know”.

To verify the data distribution, skewness and kurtosis values were divided by the respective standard error, using a significance criterion of higher than 1.96 (Cramer & Howitt, 2004). All school grades demonstrated significant negative skewness: 2nd (-3.18), 3rd (-3.43), 4th (-5.35), and

distribution was found only in 4th (2.04) and 5th grades (2.03), thus showing a more uniform layout of data than the 2nd and 3rd grades. These statistical significances were confirmed using the Shapiro–Wilk normality test.

Standardization

Table 3 shows the norms for Forms A (2nd grade) and B (5th grades (-5.70). A significant platykurtic

and B (3rd-5th) of the EACOL. The scores of the 4th and 5th grades did not differ numerically, and so these groups were combined.

Table 3

Percentile norms and classification for raw EACOL scores by school grade

School grade			Percentile	Classification
2 nd	3 rd	4 th -5 th		
0-15	0-23	0-26	0-25%	Poor reader
16-23	24-32	27-34	26-50%	not-so-good -
24-26	33-39	35-39	51-75%	not-so-good +
27-28	40-42	40-42	76-100%	Good reader

Source: own work

Discussion

By assessing the EACOL in Brazil, the present study provides information that can be of use in developing an effective tool that is relevant to education policymakers, teachers, principals, parents, and pupils. Researchers, as external advisers, can play a pivotal role as catalysts for positive actions or informed reflections by these educational stakeholders. We hope to stimulate teachers to carry out systematic evaluations of their students in elementary school, which, as the evidence shows, is an important way to prevent reading failure.

This final version of EACOL could be easily adapted to other countries, especially those that struggle with teaching Portuguese language, for instance, those with low number of people aged 15 and over that can read and write: Guinea-Bissau (55.3%), Mozambique (56.1%), East Timor (58.3%), São Tomé and Príncipe (69.5%) and Angola (70.4%) (Central Intelligence Agency [CIA], 2014). In other

nations of the Community of Portuguese-Speaking Countries, where literacy is above 90%, EACOL can be useful to screen children with risk of dyslexia; these places include Portugal and Cape Verde.

The new format of the EACOL significantly reduced the number of items in Form A (from 23 to 14) and Form B (from 27 to 21) without losing its validity. This should make the scale more attractive to the teacher, since it is now shorter and faster to complete. Even with the new modifications, however, particularly with the addition of the answer “I do not know,” some scales (although just 1%) were returned incomplete, reinforcing the conception that this problem may be due to some characteristic of the sample itself and not a failure of the scale. One theory is that the teachers in our sample prefer to decline answering an item instead of admitting that they do not know about some aspect of their student’s reading performance. One way to minimize such behavior could be to add to the EACOL’s instructions the following statement “Please always answer ‘I do not know’ in case of doubt; do not answer randomly or leave an item unanswered”.

For evidence of concurrent validity (external validation), as EACOL incorporates items that concern with accuracy in word recognition, reading speed, prosody, comprehension and the capacity for synthesis, the good correlations found with the General Reading Composite ($r = 0.737$ and 0.688) can be considered the most important result of the current study, attesting that the teachers, when provided with sound criteria, can come to reliable evaluations of their students’ reading ability.

Unlike Cogo-Moreira et al. (2012), this study found significant correlations between the EACOL, the CPM, and the SDQ. Cogo-Moreira et al. considered that the latter two measures would provide to EACOL discriminant validity. Although the CPM is sometimes referred to as a non-verbal test, it requires language to process the information, and thus is better defined as a test of general cognitive ability. Hence, a small-to-moderate positive correlation between the reading ability of the child and the CPM score is

expected (Carver, 1990). Concerning the child’s psychiatric characteristics, as assessed by the SDQ, a small but significant negative correlation is also expected. Maughan and Carroll (2006) note that disruptive behaviors impede reading progress and also the reverse: reading failure exacerbates risk for behavior problems. Thus, unlike Cogo-Moreira et al. (2012), we argue that although the variables measured by CPM and SDQ have distinct theoretical construct domains, they are not independent from each other.

As the correlations of the EACOL with general cognitive ability and psychiatric symptoms ranged from small to moderate, it is important to consider whether the teacher is taking these domains into account in her/ his evaluations of children’s reading. One way to do so is to compare these correlations with those between CPM and SDQ within the General Reading Composite. First, as the correlations between the CPM and the General Reading Composite were smaller than those with the EACOL (0.09 reduction in the value of r), we might argue that teachers can distinguish children’s general cognitive ability on the basis of their reading ability. On the other hand, the SDQ had a bigger correlation with the EACOL than with the General Reading Composite (an additional 0.12). Although small, this correlation indicates that the teacher takes the child’s behavior into consideration in his or her judgment.

As the scale was not designed to address children with excellent reading performance, an increase in the number of children in the “good” ability category occurred. This is demonstrated, for example, by the significant negative skewness distribution in all grades. On the other hand, given the numerically wide range of scores, the EACOL is an effective scale to screen for poor readers, who should in any case be the first focus for early educational interventions in schools. The strong concordance between the reading task and the EACOL of those with poor ability is in agreement with the literature, which has shown that teachers are more accurate in the assessment of poor readers, identifying 89% of children with this

type of performance (e.g., Capellini, Tonelotto, & Ciasca, 2004).

The EACOL was envisaged to offer the teachers a set of valid criterion to evaluate their pupils' reading ability in response to the demand of different researchers, as for instance, in the case of those who specifically need a sample of poor readers for an experimental study. However, the scale can also have a practical use in the school. It can be implemented as a means of establishing a comparison between the judgment of the teachers about their students' reading performance and their real achievement in the formal evaluations carried out as part of the curriculum. Any mismatch between the expected and effective achievement could lead teachers to develop a more accurate/realistic perception about the reading ability of their students. It could also alert the teachers about the aspects of their students reading that should deserve more attention.

Conclusion

Reading ability is one of the most important competences in the modern world, essential to educational, professional, and social achievements. For this reason, it is of utmost relevance to create and/or adapt scientific validated instruments for early detection of poor reading skills and risk of dyslexia. With this purpose in mind, the EACOL was developed to be a quick and efficient instrument to guide educational stakeholders in assessing the Reading Aloud (speed and accuracy in word recognition, prosody and comprehension) and the Silent Reading (text comprehension and synthesis) of elementary-school children. Furthermore, this instrument can be adapted to other countries with Portuguese as the official language or to other orthographies.

References

- Angelini, A. L., Alves, I. C. B., Custódio, E. M., & Duarte, W. F. (1999). *Manual. Matrizes Progressivas Coloridas de Raven* [Manual: Raven's Coloured Progressive Matrices]. São Paulo: Casa do Psicólogo.
- Bennett, R. E., Gottesman, R. L., Rock, D. A., & Cerullo, F. (1993). Influence of behavior perceptions and gender on teachers' judgments of students' academic skill. *Journal of Educational Psychology*, 85 (2), 347-356. <http://dx.doi.org/10.1037/0022-0663.85.2.347>
- Capellini, S. A., Oliveira, A. M., & Cuetos, F. (2012). *PROLEC: Provas de avaliação dos processos de leitura* [Reading Processes Assessment Battery]. 2 ed. São Paulo: Casa do Psicólogo.
- Capellini, S. A., Tonelotto, J. M. F., & Ciasca, S. M. (2004). Learning performance measures: formal evaluation and teachers' opinion. *Revista Estudos de Psicologia, PUC-Campinas*, 21, 79-90. <http://dx.doi.org/10.1590/S0103-166X2004000200006>
- Carreira-Perpiñán, M. Á. (1997). *A review of dimension reduction techniques*. Sheffield: Department of Computer Science, University of Sheffield. <http://dx.doi.org/10.1.1.212.994>
- Carver, R. P. (1990). Intelligence and reading ability in grades 2-12. *Intelligence*, 14 (4), 449-455. [http://dx.doi.org/10.1016/S0160-2896\(05\)80014-5](http://dx.doi.org/10.1016/S0160-2896(05)80014-5)
- Central Intelligence Agency (CIA) (2014). *The world factbook*. Washington, DC: Author.
- Cogo-Moreira, H., Ploubidis, G., Brandão de Ávila, C.R., Mari, J. de J., & Pinheiro, A.M.V. (2012). EACOL (Scale of Evaluation of Reading Competency by the Teacher): Evidence of concurrent and discriminant validity. *Neuropsychiatric Diseases and Treatment*, 8, 443-454. <http://dx.doi.org/10.2147/NDT.S36196>
- Cramer, D., & Howitt, D. (2004). *The Sage Dictionary of Statistics: A practical resource for students in the social sciences*. Thousand Oaks: Sage.
- Cury, C. R., & Golfeto, J. H. (2003). Strengths and Difficulties Questionnaire (SDQ): a study of school children in Ribeirão Preto. *Revista Brasileira de Psiquiatria*, 25 (3),

- 139-145. <http://dx.doi.org/10.1590/S1516-44462003000300005>
- Emmerich, W., Enright, M. K., Rock, D. A., & Tucker, C. (1991). *The development, investigation, and evaluation of new item types for the GRE analytical measure*. Princeton: Educational Testing Service.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38, 581-586. <http://dx.doi.org/10.1111/j.1469-7610.1997.tb01545.x>
- Goodman, R. (2005). *Questionário de Capacidades e Dificuldades (SDQ-Port)* [Strengths and Difficulties Questionnaire (SDQ-Port)]. Retrieved from <http://www.sdqinfo.com>.
- Instituto Paulo Montenegro (2011). *INAF Brasil 2011, Indicador de Alfabetismo Funcional. Principais resultados* [INAF Brazil 2011, Functional Literacy Indicator. Main results]. São Paulo: IBOPE Inteligência.
- Lúcio, P. S., & Pinheiro, A. M. V. (2013). Escala da Avaliação da Competência da Leitura pelo Professor (EACOL): evidências de validade de critério. *Temas em Psicologia*, 21, 499-511.
- Maughan, B., & Carroll, J. M. (2006). Literacy and mental disorders. *Current Opinion in Psychiatry*, 19, 350-355. <http://dx.doi.org/10.1097/01.yco.0000228752.79990.41>
- Pasquali, L., Wechsler, S., & Bensusan, E. (2002). Matrizes Progressivas do Raven Infantil: um estudo de validação para o Brasil [Raven's Colored Progressive Matrices for Children: a validation study for Brazil]. *Avaliação psicológica*, 1(2), 95-110.
- Pinheiro, A. M. V. (2013). *Prova de Leitura e de Escrita de palavras [Reading and Writing words test]*. Relatório Técnico Final aprovado pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais —FAPEMIG (FAPEMIG). Número do processo: APQ-01914-09.
- Pinheiro, A. M. V. (2013). *Prova de Leitura de palavras [Word Reading Test]*. Relatório Técnico Final aprovado pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais —FAPEMIG (FAPEMIG). Número do processo: APQ-01914-09.
- Saur, A. M., & Loureiro, S. R. (2012). Psychometric properties of the Strengths and Difficulties Questionnaire: A literature review. *Estudos de psicologia (Campinas)*, 29 (4), 619-629. <http://dx.doi.org/10.1590/S0103-166X2012000400016>
- Snowling, M. J., Duff, F., Petrou, A., Schiffeldrin, J., & Bailey, A. M. (2011). Identification of children at risk of dyslexia: The validity of teacher judgements using “Phonic Phases.” *Journal of Research in Reading*, 34, 157-170. <http://dx.doi.org/10.1111/j.1467-9817.2011.01492.x>
- Soares, T. M., Fernandes, N. S., Ferraz, M. S. B., & Riami, J. L. R. (2010). A expectativa do professor e o desempenho dos alunos [Teacher's Expectation and Students' Performance]. *Psicologia: Teoria e Pesquisa*, 26(1), 157-170.
- Todos pela Educação (2013). Instituto Paulo Montenegro/IBOPE, Fund. Cesgranrio, Inep. Retrieved from <http://www.todospelaeducacao.org.br>
- Unesco (2007). *Literacy Initiative For Empowerment 2006-2015. Vision and Strategy Paper*, 3rd edition. Hamburg: Unesco Institute for Lifelong Learning.
- Verbic, S. (2012). Information value of multiple response questions. *Psihologija*, 45, 467-485. <http://dx.doi.org/10.2298/PSI1204467V>
- Vilhena, D. A., Sucena, A., Castro, S. L., & Pinheiro, A. M. (2016). Reading Test-Sentence Comprehension: An Adapted Version of Lobrot's Lecture 3 Test for Brazilian Portuguese. *Dyslexia*, 22 (1), 47-63. doi: 10.1002/dys.1521

Appendix

Items and scoring rubric for the EACOL (Scale of Evaluation of Reading Competence by the Teacher)

Item	True	False	Sometimes	I do not know
1. <u>Reads but cannot tell what was read, even when stimulated with questions.</u>	0	2	1	1
2. <u>Reads with intonation compatible with the punctuation marks, expressing emotions and feelings according to the text read. For example, gives an intonation of questioning in the whole sentence when there is a question mark in the text. Gives intonation of joy or surprise in the whole sentence when there is an exclamation mark.</u>	2	0	1	1
3. <u>Reads very slowly, without rhythm, spelling out each syllable; does not observe the punctuation marks.</u>	0	2	1	1
4. Reads by spelling out both "new" and "known" words.	0	2	1	1
5. <u>Does not take into account the intonation compatible with the punctuation marks, reading in a monotone manner.</u>	0	2	1	1
6. Reads "new" words correctly.	2	0	1	1
7. Reads clearly, without "stumbling" or "swallowing" syllables. Someone who hears can understand what is being read.	2	0	1	1
8. Frequently makes mistakes when reading "new" words.	0	2	1	1
9. Quickly and correctly reads both "known" and infrequent words.	2	0	1	1
10. <u>Seems to have understood what was read when asked about the text read.</u>	2	0	1	1
11. Has great difficulty in reading aloud.	0	2	1	1
12. <u>Reads with rhythm, neither too slowly nor too fast.</u>	2	0	1	1
13. <u>Can identify characters, places, and main ideas after the first reading.</u>	2	0	1	1
14. Reads without pronouncing words or without moving the lips, only moving the eyes.	2	0	1	1
15. <u>Not able to orally summarize what was read.</u>	0	2	1	1
16. <u>Is able to identify the subject from the title and vice versa.</u>	2	0	1	1
17. Cannot read without movements of the lips or without pronouncing the words.	0	2	1	1
18. <u>Is able to choose a title for passages with no title or even give an alternate title for titled passages.</u>	2	0	1	1
19. Does not identify characters, places, or main ideas. *Corresponds to item 5 in Form A, which was excluded.	0	2	1	1
20. Can orally summarize the text read.	2	0	1	1
21. <u>Does not identify the subject from the title or vice versa.</u>	0	2	1	1
x. <u>Reads with difficulty the "known" words.</u> *Item present only in Form A.	0	2	1	1

Form A (2nd grade) contains only the underlined sentences, while Form B (3rd–5th grade) contains both underlined and non-underlined sentences. Each item is followed by the possible responses: "True," "False," "Sometimes," and "I do not know."

Notes

- * Research article. This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, grant No. 134357/2013-2), and had no involvement in the study design; collection, analysis and interpretation of data; writing of the report; or decision to submit the article for publication.