Journal of Psychological and
Educational Research
JPER - 2018, 26 (2), November, 67-88

DEVELOPMENT AND VALIDATION OF A TEACHER RATING SCALE PRIMARY SCHOOL CHILDREN WITH READING DISABILITY

Layes Smail *
Normandie University, France
University of El-Oued, Algeria

Mamadi Chaouki
University of El-Oued, Algeria

Abstract
This study was designed to evaluate the internal structure and discriminant evidence of a teacher rating scale for screening children with reading disability. A 16 items teacher rating scale was developed to provide a brief screening measure for reading disability. The scale ratings were obtained from teachers (n=8) of 280 Arabic speaking children divided into two groups: typical readers (n=194), and disabled readers (n=86). Exploratory and Confirmatory Factor Analyses revealed four correlated but separable dimensions that were labelled word recognition, reading comprehension, writing, and memory. The Factorial Confirmatory Analysis (FCA) provided a reasonable fit to the data. Children rated as having severe scores in the scale, gained lower scores on objective measures of reading decoding and comprehension skills. In addition, the sensitivity of the scale was acceptable. These results suggest that the developed scale may provide a useful method to screen for reading disability in Arabic children in primary school.

Keywords: reading disability; teacher rating scale; screening; primary school

Introduction

Reading disability (RD) is a neurodevelopmental disorder which primarily affects the development of accurate and/or fluent word recognition,

Correspondence concerning this paper should be addressed to:

* Ph.D., Professor, Research Center on Psychological Functioning and Dysfunctioning, Normandie University, Rouen, France. Department of Psychology, University of El-Oued, El-Oued, Algeria. Address: Chott, El-Oued 3900. E-mail: *Smail.layes@mail.com*

spelling and decoding abilities (APA, 2011; IDA, 2002). Secondary consequences may include problems in reading comprehension and reduced reading experience that can hinder growth of vocabulary. (Lyon, Shaywitz, & Shaywitz, 2003). According to Thambirajah (2011), essential features of RD include inaccurate and slow reading characterised by distortions, omissions and substitution of words or parts of words and poor spelling. Prior studies also showed that Arabic individuals with RD are weak at multiple reading related abilities, including reading speed (Layes et al., 2014), reading comprehension (Layes et al., 2015a), phonological and morphological awareness and working memory (Layes et al., 2015b), and visual-orthographic skills with the first three being their major difficulties.

The high prevalence of RD and its frequent co-occurrence with other disorders, suggests that RD assessment measures should be included systematically in clinical assessment in school. The importance of early intervention for RD has highlighted the need for reliable and valid screening instruments. However, a full RD assessment requires the administration of time consuming standardized tests of learning achievement and cognitive ability by trained examiner. Questionnaires and rating scales provide a time-saving method of evaluating RD indicators, and offer a cost-effective and valid technique of obtaining information on children's learning behaviour across settings (Dillman, 2000). Thus, a two-stage assessment process is strongly recommended, including the implementation of an initial screening tool and a follow-up complete evaluation (Ho, 2010).

Over the years, considerable evidence has been accumulated supporting the effectiveness of the use of behavioural checklists to assess developing various kinds of disabilities in children (Wagner, 2003). These scales require parents and teachers to indicate the presence of some general behaviours or the frequency and intensity of these behaviours of the children (Chan, Ho, Tsang, Lee, & Chung, 2004; Mash & Terdal, 1997). Although behavioural rating scales offer many benefits to the assessment process including administration convenience (Shelton & Barkley, 1994), there is still a lack of screening tools available for RD, which calls for experimental investigations into this issue. The rating scale developed in the present study could be a useful tool to identify Arabic children with reading disability and therefore may help in overcoming academic difficulties in primary school.

A prominent example is the Dyslexia Screening Instrument, which consists of 33 statements aiming to identifying students who display behaviour significantly associated with reading, spelling, writing, or language processing difficulties. Another noteworthy screening tool designed to measure risk of reading disability in school-age children is The Colorado Learning Disabilities Questionnaire - Reading Subscale (CLDQ-R) (Willcutt, Boada, Riddle, Chhabildas, DeFries, & Pennington, 2011). Normative scores for this questionnaire were developed based on parent-reports of their 6-18 years old children, as well as actual reading testing of these children.

Using scales based on students' manifestations associated with RD could be an efficient alternative for initial screening. Therefore, an assessment tool is needed to capture the individual reading difficulty and to determine if the student displays a pattern of performance that suggests the presence of reading disability including secondary issues, such as problems with reading comprehension due to poor decoding, or problems in written expression due primarily to spelling difficulties (Mather & Wendling, 2012).

Nevertheless, one obstacle to the identification of RD in schools is a lack of teacher awareness about perceptions of identification. According to Wadlington et al. (1996), dyslexia is still a confusing term for most teachers and educational professionals, and they find difficulty in differentiating between dyslexia and other relating difficulties. For this purpose, teachers need to be made sensitive for screening tools of children and take appropriate measures to overcome the problems raised in the critical period for acquiring the skills of reading and writing corresponding to 6 to 8 years of age (Joshi & Vanaja, 2016). The identification of RD is of vital importance for school children otherwise it will have a devastating effect on the children's life and achievement (Bailet, Repper, Piasta, & Murphy, 2009; Vellutino, Scanlon, Small, & Fanuele, 2006).

Despite the fact that children may experience reading difficulty after the initial stages of literacy instruction, there is a paucity of literature examining teacher screening of reading disabilities and methods of identifying these children. In the literature, most researches about reading disabilities evaluation focused on developing self-report questionnaires in adults (e.g., Gilger, 1992; Harrison & Nicols, 2005; McKenzie & Paxton, 2006; Pennington & Lefly, 2001; Schulte-Körne, Deimel, & Remschmidt, 1997; Snowling et al., 2012); and learning disabilities screening tools in school age children (e.g. Lange &

Thompson, 2006; McKenzie, Megson, & Paxton, 2008; NCLD, 2007). Thus, a cost-effective and valid screening instrument for RD that can be easily administered by teachers would constitute a crucial contribution to the enrichment of the assessment equipment in educational settings.

The present study investigated whether a teacher rating scale could provide a valid measure of reading ability which could be useful for the identification of RD in primary school children. To our knowledge, there is no validated scale designed to screen RD in Arabic primary school children. The primary reason for validating the rating scale was for use in school environment investigating for reading disability. In this work, we describe the validation of a reading disability teacher-report rating scale that may provide a useful screening instrument for use in clinical as well as in research studies. The current study focussed on the participation of teachers in identification of RD, that is, to design an instrument that would enable to detect basic reading disabilities that are important to them to be recognised and measured.

Objective

The purpose of this study was to (a) provide evidence to support the reliability and validity of the developed scale as a teacher screening tool for Arabic children with RD; and (b) to explore the relationships between the reading-related characteristics assessed by the scale and the profiles of reading, comprehension skills measured by external tests in Arabic children readers, and (c) to investigate whether literacy characteristics measured by the scale could differentiate children with RD from those without RD.

Method

Participants

Initially, 300 native Arabic speaking children from 4th and 5th grade primary school were considered for the study. Candidates for the reading disabled and the control groups were excluded if they had any abnormalities in vision, even corrected with normal visual acuity, or had any abnormalities on clinical examination. 20 children with previous history of neurological disease, impaired sensory-motor coordination, social deprivation and inadequate schooling, were excluded from the study. 280 children were retained for the

final study. All children had normal intelligence based on Raven's Standard Progressive Matrices.

The diagnostic criteria of RD was based on the standardization of the sample mean in reading test, below (-1) SD (4th ed.; DSM-IV; American Psychiatric Association, 1994). For the reading disabled group, children were required to have a diagnosis of RD without having received treatment or psycho-pedagogical support for literacy on the basis of school health reports. The sample was composed of a group of typical readers, age mean 126.36 months (SD=7, 07) and a group of disabled readers, age mean 126.12 months (SD=7, 25).

Materials

Reading achievement measures

This test was elaborated beforehand for the purpose of assessing reading skills in children ages 9-11 (Layes, Lalonde, & Rebaï, 2014; Layes, Lalonde, Mecheri, & Rebaï, 2015). A set of 80 stimuli was given (40 frequent words and 40 infrequent words). The words used varied in length (mono-syllabic and disyllabic) and frequency (high and low); in addition to a list of 20 pseudo words which varied in orthographic length. Participants were required to read the stimuli aloud without any limitation in time. The number of correct responses (accuracy) and reading speed (time in seconds per list) were scored. Inter-consistency coefficient for word and pseudo-word reading was high (α=.91).

Single word reading test. Participants read 80 partially vowelized words, 40 words matched for frequency (40 frequent/40 infrequent words). Each list was presented on a white sheet of A4 paper. Participants were required to read aloud the items correctly.

Pseudo word identification task. Pseudo word identification is a trustworthy measure of phonological coding and grapheme-phoneme decoding skills (Pugh et al., 2013) and represents the most commonly used task in transparent systems (Abu-Rabia & Abu-Rahmoun, 2012). This task consisted of a list of 20 partial vowelized pseudo-words varying in orthographic length (CVCV/CVCVC).

Reading comprehension. This test (Layes et al., 2015b), is a modified version of the reading sentence comprehension subtest of Khomsi's (1999) Reading Words and Comprehension-Revised test (LMC-R), composed of 15

items. The sentences were in some way designed in ascending order of syntactic difficulty (structural length). Reading comprehension included sentences along with pictures. Children silently read a sentence and selected the target picture from four choices best representing the sentence without any limitations on time. The other pictures were distracters that included only some information from the sentence. This reliance on pictures as a source for extracting meaning parallels RC instruction received in primary school. It should be noticed that the use of pictures (rather than oral questions) may reduce the involvement of any listening comprehension load constraining RC in poor-reading children, known to have poor (Nation, 2005).

The teacher rating scale for reading

Based on models of reading development, measures of word decoding and recognition and reading comprehension are prime candidates for screening measures designed to discriminate good and poor readers (Shankweiler et al., 1999). The instrument is a 16 item teacher-report rating scale of student reading-related characteristics that can be observed in classrooms by teachers, that provides a brief screening of the child's functioning within several reading domains: reading words and decoding (6 items, for example: "there are great difficulties to learn the association between letters and sounds"); reading comprehension (4 items, for example: "there are evident problems to understand the meanings of words and short sentences"); writing (4 items, for example: "the child's writing letters and words are usually unrecognizable"), and memory (2 items, for example: "there is great difficulty in the repetition of orally presented words"), since these symptoms often occur together in people with reading disabled children and are sufficiently common to be regarded as dyslexia associated traits (DSM-4).

Screening scale was given to the teachers of 4^{th} and 5^{th} classes. They were briefed about the RD of the children, the purposes and nature of the screening and the filling of the screening scale. They were further asked to report the real reflection of their observation about the children reading abilities. Teachers were asked to respond to each item (e.g. "he/she reads very slowly") on a 4-point scale ranging from a rating of 0 (not at all) to 4 (a great deal). Higher scores indicate greater levels of perceived reading difficulty. However, the tasks of word recognition, pseudo word identification, reading comprehension were conducted by trained school psychologists.

Statistical analysis

To assess the number of dimensions of RD assessed by the developed rating scale, initial Exploratory Factor Analysis (EFA) was applied to the 22 items using SPSS (version 22), and a subsequent Confirmatory Factor Analysis (CFA) was used to test whether the factor structure could be adequate and if the hypothesized model was a good fit to the observed data. We hypothesized that these analyses would identify separable dimensions of reading, comprehension, writing and memory. Evidence of the predictive validity of scale comes from its relationship to measured reading and writing skills (Elbro, Nielsen & Petersen, 1994). Internal consistency (Cronbach's alpha) was calculated for the both reading scale and sub-scales.

In order to build a predictive model for group membership, a discriminant function analysis was performed to ensure if the predictor variables included in the scale rating provide the best discrimination between the two groups of readers. Furthermore, discriminative validity was also examined using Area Under the Curve (AUC) scores from non-parametric Receiver Operating Curve (ROC) analyses, which is a plot of true positive Vs. false positive results. Swets (1988) suggested the following benchmarks for interpreting AUC scores: 0.50-0.70 (low accuracy), 0.70-0.90 (moderate accuracy), and > 0.90 (high accuracy).

Results

The present study was aimed to develop a teacher rating scale to screen reading disabled children and investigate its efficacy to identify the children with reading disability. The developed scale with 22 questions was administered on all participants. This instrument was prepared for teachers based on common observation of the children with RD. The screening scale consisted of basic assumptions which were true for the identification of RD and four factors termed as word recognition and decoding, reading comprehension, writing and memory factors. The scores obtained were subjected to item analysis revealing adequate reliability statistics of both Cronbach's alpha (.93) and Spearman-Brown Coefficient (.86) associated with the items.

Dimensions of reading characteristics in the scale

Construct validity was determined by means of Exploratory Factor Analysis (EFA) using descriptive statistics, principal axis factoring extraction

method and direct Oblimin rotation, as this method is preferred when there are strong theoretical grounds to consider that factors might correlate (Field, 2007). The various indicators of factorability were good: the Kaiser-Mayer-Olkin (KMO) measure of sampling adequacy was .90; it indicated that the correlation matrix was indeed suitable for factoring. Further, Bartlett's test of sphericity was highly significant (χ^2=3410.305, *df*=136, *p*<.001), which indicated that correlations between items were sufficiently large, and the residuals indicated a good solution. As this analysis was intended to validate pre-defined components, number of component was restricted to 4 (Component 1=reading and decoding; Component 2=reading comprehension; Component 3=writing; Component 4=memory). Together, the four factors emerged with an eigenvalue greater than 1, explain 79,36% of the variance. Inter-item correlation ranges from 0.28 to 0.66, which means items are reasonably correlated and contribute to same construct.

Out of 22 items, 6 items were deleted as these items do not statistically match and fall under the factors as in original scale. After deletion of poor items, finalized factor analysis yields 16 items. Negative factor loadings of items with factors are observed indicating that these items are opposite of what positive factor loading items are measuring. In that case, negative factor loadings are ignored as they represent to measure opposite of what a certain factor measures. Table 1 represents factor loadings along with Eigen values, percentages of variance, and cumulative percentages of variance.

Table 1. The proportion of total variance explained by the initial solution

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	9,52	56,03	56,03	9,27	54,57	54,57	3,73	21,99	21,99
2	1,89	11,123	67,15	1,651	9,71	64,29	3,14	18,52	40,51
3	1,23	7,26	74,42	,98	5,78	70,07	2,87	16,91	57,43
4	,84	4,94	79,36	,60	3,53	73,61	2,75	16,18	73,61
5	,69	4,09	83,45						
6	,50	2,95	86,41						
7	,43	2,53	88,95						
8	,32	1,89	90,84						
9	,30	1,76	92,61						
10	,27	1,60	94,21						
11	,22	1,30	95,51						

Table 1. The proportion of total variance explained by the initial solution - *continued*

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
12	,17	1,00	96,52						
13	,15	,93	97,46						
14	,12	,73	98,19						
15	,11	,68	98,88						
16	,08	,51	100,0						

The reading and decoding subscale appears most robust of the four, containing the first six (6) items, all of which loaded highly (≥.40) on the factor. The reading comprehension contained four items (7, 8, 9, 10), the item-factor loadings were high (≥.48). The writing scale (Items 11, 12, 13, 14) all of them reach an adequate factor loading (≥.40) according to accepted scale development standards (De Vellis, 2012). For Memory containing only two items (15 and 16) but it is clear conceptually and item-factor loadings were high (≥.48). Therefore, factorial validity of the screening scale was well established.

Table 2. Factor loading for EFA with Oblique Rotation (Direct Oblmin) of reading disabilities scale (N=280)

Factor loading			
Reading	Comprehension	Writing	Memory
,763	,705	,873	,592
,752	,854	,547	,640
,555	,862	,855	
,602	,951	,847	
,486			
,848			

Internal consistency reliability

Scale internal consistency reliability was determined by calculating Cronbach coefficient alphas (Table 3). The Cronbach Alpha of the entire scale was .91. The internal consistency of each sub-scale was also assessed using Cronbach alpha, indicating a high reliability among all subscales, for reading (α=.88), reading comprehension (α=.90), writing (α=.92), and fairly adequate for memory (α=.67). Alpha coefficients for total and subscales in Table (3)

range from .67 to .92 are far above the minimum level of .65 Cronbach's alpha level. Inter-subscale and total-subscale correlations are also statistically significant (all *P*s<.01).

Table 3. Alpha coefficients and correlations between reading, comprehension, writing, memory and total score of the scale

Variables	N° of item	α	M	SD	Reading	Comp	Writing	Memory	Total score
Reading	06	.88	16.26	4.58	-	.65**	.61**	.45**	.88**
Compreh.	04	.90	11.28	3.22		-	.54**	.63**	.84**
Writing	04	.92	9.07	4.02			-	.45**	.83**
Memory	02	.67	5.12	1.58				-	.67**

Note: ** p<.01

Moreover, Split Half Reliability of the scale was calculated. (r=.81; Part I, r=.86; Part II, r=.82). Thus, the final scale developed included 16 items with four domains including reading and decoding skills (6 questions), reading comprehension (4 questions), writing (4 questions), and memory (2 questions), the scores for the scale range between 0-48.

Table 4. Means and standard deviations of RD among participants

Variables	With RD (n=194)	Without RD (n=86)			95% CI		
	M (SD)	M (SD)	t(1.96)	P	UL	LL	Cohen's *d*
Reading	15,00 (4,50)	19,10 (3,35)	8,46	,000	-3,14	-5,06	1.03
Compreh.	10,55 (3,21)	12,93 (2,60)	6,02	,000	-1,59	-3,14	0,81
Writing	8,18 (3,72)	11,10 (3,97)	5,93	,000	-1,95	-3,89	0,75
Memory	4,82 (1,52)	5,77 (1,53)	4,79	,000	-,55	-1,3	0,62
Total	38,56 (10,82)	48,91 (8,34)	8,70	,000	-8,00	-12,69	1,07

Note: CI= Confidence Interval; UL= Upper Limit; LL= Lower Limit

Independent sample t-test and Cohen's *d* in (Table 4) also reflect that scale clearly differentiates between reading disabled and non-reading disabled participants. Group with RD significantly obtains high scores than their peers without RD.

Construct validity

We began by examining the predictive validity of the four factors in relation to measures of literacy. We expected the reading scale score to

correlate strongly with measured reading and spelling skills. Table 5 shows the pattern of correlations between the scale scores and raw scores on measures of word and non-word reading, comprehension, writing and memory. All coefficients are statistically significant, and the pattern of correlations is important between all measures. The highest coefficients correlate reading and comprehension sub-scales with word reading accuracy and speed and pseudo-words decoding, suggesting that reading word is as sensitive measure as reading comprehension. In contrast, correlations between the writing and memory subscales and word reading were moderate, but stronger with pseudo word decoding.

Table 5. Correlations between the scale scores (at the top), and measures of word, pseudo word reading, and reading comprehension

Variables	Reading	Compreh.	Writing	Memory	Total scale
Word read Accuracy	-.53***	-.45***	-.34***	-.26***	-.50***
Word read Speed	.44***	.38***	.23***	.24***	-.41***
Pseudo Word	-.53***	-.53***	-.37***	-.31***	-.55***
Read Comprehension	-.25***	-.25***	-.19***	-.18***	-.27***

To investigate whether teachers' ratings with the scale could effectively differentiate students with RD from those without RD as defined by test scores of reading words, we conducted a multivariate analysis of variance on the four subscales' scores of reading characteristics, reading comprehension, writing and memory as dependent variables using the group of readers as the grouping variable. The results showed significant differences for the four scores of reading-related subscales, Pillai's trace=.189, $F(4, 27)=16.03$, $p<.001$, $\eta^2=.18$. Follow-up univariate ANOVAs indicated that both reading and decoding, reading comprehension, writing and memory were significantly different for groups of readers respectively, $F(1, 27)=57.30$, $p<.001$; $F(1, 27)=36.26$, $p<.001$; $F(1, 27)=35.22$, $p<.001$; and $F(1, 27)=22.95$, $p<.001$.

A discriminant analysis using the whole scale to predict membership in the group with RD was conducted. The overall results were significant (Wilks $\lambda=.73$, $\chi^2(1, n=280)=84.80$, $p<.001$), indicating that the scale significantly differentiated children with RD from those without RD. With prior probabilities based on group sizes for classification, the discriminant function correctly

classified 76.8% of the 280 students. The large eigenvalue value (.36) indicates that the variance in the dependent variable is explained by that function.

To determine a scale cut-off value, diagnosed reading disabled and non-disabled groups were examined using a Receiver Operating Characteristic (ROC) curve analysis, employing the AUC (area under the curve) as the index of discriminant ability. The ROC curve (Figure 1) follows the left-hand side and then the top border, showing that mean reading ratings on the scale effectively discriminated children with RD, indicating a high classification accuracy of scale scores as indicated by achievement scores, from those without RD. Specifically, the area under the curve (AUC) was .84 (95% CI=.78, .90, p<.001). A cut-off score of 50, that is, we determined that classifying individuals with scale scores greater than 50 as reading disabled, resulted in the best balance between sensitivity (94%) and 1 - specificity (55%). This cut-off also resulted in a reasonably good specificity of in correctly identifying children without RD.

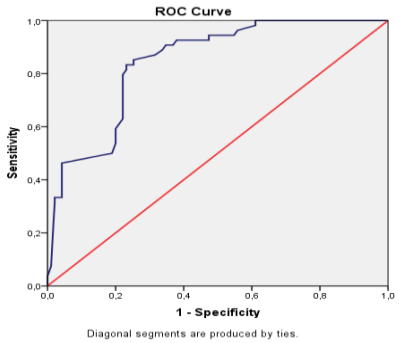

Figure 1. The ROC curve corresponding to the data set of the RD scale for teacher

Confirmatory factor analysis was used to confirm the factor structure of the scale (Figure 2). For CFA, Hoyle (2000) recommended using fit statistics that possess different computational logic. Thus, absolute fit was assessed using the Standardized Root Mean Square Residual (SRMR) and the Root Mean Square Error of Approximation (RMSEA); and comparative fit was examined using Bentler's Comparative Fit Index (CFI). Fit indices for the 16 item scale were: $\chi^2(98)=339.46$, p<.001. Although the chi-square is significant, many researchers disregard this index if both the sample size exceeds 200 or so and

other indices indicate the model is acceptable. The relative chi-square index, also called the normed chi-square, which the value equals the chi-square index divided by the degrees of freedom, might be less sensitive to sample size. The criterion for acceptance varies from less than 2 (Ullman, 2001) to less than 5 (Schumacker & Lomax, 2004). In our study this index is appropriate (339.46/98=3, 46). Furthermore, RMR=.05 (less than .08); RMSEA=.09 (>.06); GFI=.88; CFI=.92; NFI=.90. The last values meet the criteria (.90 or larger) for acceptable model fit. Overall, the CFA fit indices indicated that the model is acceptable.

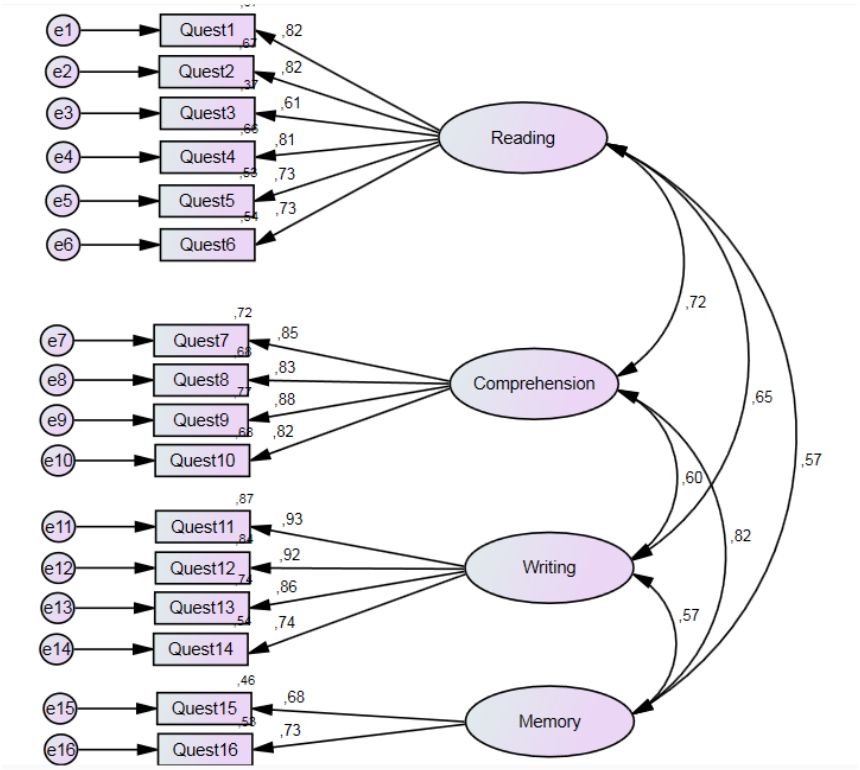

Figure 2. CFA diagram of the 16 items reading scale for teacher

Discussion

In order to make effective helpful provision to children with reading disability, teachers and related professionals need an effective measure to

identify these cases. As literacy difficulties and specific cognitive deficits related to reading could be readily detected at the behavioural level (Frith, 1997), we developed the RD rating scale. On the basis of this instrument, teachers could obtain a brief profile of their students' reading-related characteristics by observing their everyday behaviours. This study describes the development and validation of the scale and offers some evidence for adequate reliability and validity of the newly developed measure.

Reading, comprehension, and writing difficulties are basic component of RD. To identify these disabilities among children, screening scale was used in this study. Reading disability scale significantly differentiated children with and without reading disabilities. Present study also investigated the co-existence of various RD among students. In line with previous studies, our findings show significant comorbidity of word reading disability with reading comprehension (Wong et al., 2016), writing (Döhla & Heim, 2016) and working memory skills (Willcutt et al., 2013).

We conducted exploratory factor analyses on teachers' ratings on the scale and separate factors emerged. We constructed individual sub-scales from the scale on the basis of the items within each factor (construct), and this resulted in 4 subscales: reading decoding, reading comprehension, writing, and memory. On the basis of factor analysis, it was concluded that there was high loading in factor analysis for factors; reading words and pseudo words, reading comprehension, writing and memory factors of the screening scale.

All four factors demonstrated adequate internal consistency, with Cronbach's Alphas exceeding the normally accepted .65 limit (Goforth, 2015). All the sub-scales showed good internal consistency reliabilities. Furthermore, the significant differences observed among the two groups of readers in the expected direction for all subscales of the scale (Table 4), make them interesting for further research as well as for clinical purposes.

Convergent evidence for the reading disability scale is provided by significant correlations with standardized measures of reading achievement (overall $p<.001$). In addition, individuals who met diagnostic criteria for RD scored significantly higher on the reading sub-scale than on any other sub-scales, and the mean of the group with RD was significantly higher than the means of groups without RD. These results provide strong convergent and discriminant evidence for the reading scale. Further, the correlation analyses indicated that all four constructs of the reading scale significant correlated with

participants' performances on all reading measures (decoding accuracy and speed, reading comprehension). This finding is consistent with Salvesen and Undheim (1994), who found that the degree of correspondence between teacher assessments of low achievement and objective testing results was relatively good.

This study revealed that all four scores of reading-related characteristics of the scale could significantly differentiate children with RD from those without RD. The discriminant analysis also showed that the scale had adequate rate of correct classification (82.2%). These findings lent support to the reliability and validity of the developed scale as a screening tool for Arabic primary school students having reading disability.

The validity of the 16-item scale was tested in various ways. Comparing mean scores of the two groups of readers, we found, as expected, participants who had been diagnosed with RD scored significantly higher than relatives. Children who classified as reading disabled gained higher scores on the reading scale than those who were not. Importantly, the reading factor showed strong concurrent relationships with measured literacy skills (particularly decoding accuracy and fluency), indicating that the rated severity of individuals reading difficulties correlated with their scores on the reading scale and, consistent with previous research, negatively with measures of reading and spelling (Schulte-Körne et al., 1997; Wolff & Lundberg, 2003). Together these findings indicate that the scale provides a valid continuous measure of reading skills, and the protocol can identify reading associated traits including difficulties with reading comprehension and writing.

To determine the applicability of the reading scale as a screening tool for children with RD, we used a ROC curve analysis and determined that classifying those with scores greater than .50 as RD resulted in the best balance between sensitivity and specificity. The sensitivity and specificity of a test quantify the diagnostic ability of the test and have important clinical value. The sensitivity of a test provides information regarding its ability to identify people at risk for a diagnostic, and the specificity of a test confirm the presence of the diagnostic.

High sensitivity and specificity indices provide confidence that the screening tools will accurately identifying the students who encounter difficulty and identify students in need of intervention. An additional consideration for screening is that the screening tools are ideally brief and easy to administer and

score. Jenkins, Johnson and Hileman (2004) recommended a multivariate approach to selecting predictor variables to include not only relevant reading variables such as comprehension, word recognition, and decoding, but also other nonreading variables that have reasonable correlations with reading.

In addition, we performed a Structural Equation Modelling (SEM) diagram in order to show latent variables related to the various measures. Such variables represent the shared variance between two measured constructs. The hypothesized model (Figure 3) contained seven (07) significant paths between variables remained in the best fitting model. The interpretation of these statistics is that the resulting path model, regarding reading abilities, adequately fits the observed data from our sample of readers and our selected variables and their respective measures.

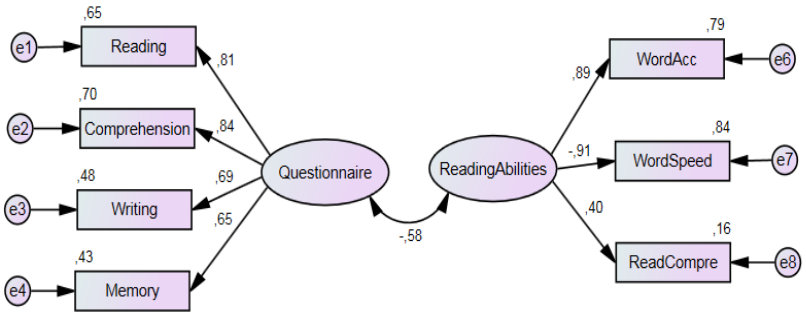

Figure 3. SEM diagram showing paths between latent variables related to the various measures

Our findings demonstrated that the teacher rating scale focus on the well-known primary reading characteristics of RD: difficulty reading words in isolation, difficulty accurately decoding unfamiliar words, difficulty with oral reading (slow, inaccurate, or laboured) difficulty spelling (Dyslexia Handbook Revised, 2014). The child shows difficulties in recognising letters, in associating graphemes to phonemes, and in automating these tasks. In students with RD, reading aloud turns out to be slower and/or less accurate than expected, in relation to age, school level, and education received. This study demonstrated that the reading and encoding sub-scale was the most correlated

with external reading measures and the strongest discriminator between reading disabled and typical readers, which is consistent with Garside and Wilkins (2002), who stated that the common reading errors made by dyslexic students at first grade are reversals reading, omissions, additions (inserting words or letters not in the text), substitutions (saying a different word from what is in the text), transpositions. Additional warning signs include: difficulty in decoding unknown words, slow, laboured oral reading, and very poor spelling.

Findings showed that reading disability scale is appropriate tool for screening children with and without RD in primary schools. These four signs of RD (decoding accuracy and speed, comprehension and writing) could be measured simultaneously and comprehensively among children by using this scale, indicating that this screening tool may help teachers and schools professionals to examine and understand the difficulties encountered in reading, comprehension, and writing domains.

Even though the scale can be completed quickly by teachers to screen efficiently for a broad range of children with reading disabilities, and can be used in clinical practice, scores from this measure do not replace diagnostic tests and is not intended to provide clinical diagnosis or to guide treatment planning in isolation. Instead, this norm-referenced rating scale provide reliable and valid indicators of areas in which an individual appears to be experiencing significant reading difficulty in comparison to others of the same age, and these areas can then be targeted directly for more intensive evaluation.

Although research on the role of memory has produced conflicting results, reporting either significant contributions of memory to reading (e.g., Oakhill, Cain, & Bryant, 2003) or not (e.g., Fowler & Swainson, 2004), our findings support the role of memory in reading development and are in agreement with previous study demonstrating that working memory resources is an important determinant of children's word reading ability and reading comprehension (Layes et al., 2015a). One argument for including working memory is based on Perfetti's (1985) verbal efficiency theory, which suggests that slow word recognition places high demand on working memory, inhibiting cognitive resources necessary for comprehension (Perfetti, 2007).

Limitation of the study and future directions

We consider the teacher rating scale a valid and reliable screening measure of children with RD. Although these preliminary results are encouraging, it will never be appropriate for clinicians to use the scale in

isolation to make categorical diagnostic or treatment decisions regarding a specific individual.

Although internal consistency of the behavioural scale is high, future examinations on its test-retest reliability are necessary. It would also be informative to carry out longitudinal studies. Children rated by teachers on the scale could receive comprehensive testing both at the time of the initial ratings and later. Long-term follow-up on children would be essential to determine whether the measure has a strong predictive value over time. Further, in this study assessment was made on basis of teachers' reports. By administering rating scale on parents may provide better, objective, and direct information could be established.

Conclusion

In conclusion, this study yielded to a development of a valid and reliable scale of children with RD, suggesting that this instrument may provide a useful screening measure for reading disabilities in both research and clinical settings. However, academic institutions may provide assistance based on scale results and investigate more detailed problems of these children having reading difficulties at the school level.

Exploratory and confirmatory factor analyses of the scale revealed four correlated characteristics of reading difficulties in children. Results provide strong convergent and discriminant evidence for scores on the 16-item reading scale, but additional research is needed to address specific weaknesses identified in each of these sub-scales. In clinical practice or screening of children in academic institutions, a cut-off providing the best balance between sensitivity and specificity must be selected. With the .50 cut-off demonstrating a sensitivity of 84.5% and specificity of 83.7%, the use of the scale in large-scale screening efforts could be justified.

References

Abu-Rabia, S., & Abu-Rahmoun, N. (2012). The Role of Phonology and Morphology in the Development of Basic Reading Skills of Dyslexic and Normal Native Arabic Readers. *Creative Education, 3,* 1259-1268.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders:* DSM-IV. 4th ed. Washington (DC): American Psychiatric Association.

Ching, B. H.-H., Ho, C. S.-H., Chan, D. W., Chung, K. K.-H., & Lo, L.-Y. (2013). Behavioural Characteristics of Chinese Adolescents with Dyslexia: The Use of Teachers' Behaviour Checklist in Hong Kong. *Applied Psycholinguistics*. http://hdl.handle.net/10722/187219

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-Risk Readers in First Grade for Early Intervention: A Two-Year Longitudinal Study of Decision Rules and Procedures. *Journal of Educational Psychology, 98,* 394-409.

Coon, K. B., Waguespack, M. M., & Polk, M. J. (1994). *Dyslexia screening instrument.* San Antonio TX: The Psychological Corporation Harcourt Brace & Company.

Dillman, D. (2000). *Constructing the questionnaire: mail and internet surveys*. New York: John Wiley & Sons.

Döhla, D., & Heim, S. (2016). Developmental Dyslexia and Dysgraphia: What can We Learn from the One about the Other? *Frontiers in Psychology, 6*, 2045. Doi: 10.3389/fpsyg.2015.02045.

Dyslexia Handbook. (2014). *Procedures concerning dyslexia and related disorders*. Texas Education Agency https://tea.texas.gov/academics/dyslexia/

Elbro, C., Nielsen, I., & Petersen, D. K. (1994). Dyslexia in Adults: Evidence for Deficits in Non-Word Reading and in the Phonological Representation of Lexical Items. *Annals of Dyslexia, 44*, 205-226.

Field, A. (2009). *Discovering Statistics using SPSS*. London: Sage.

Fielding-Barnsley, R., & Purdie, N. (2005). Teacher's Attitude to and Knowledge of Metalinguistics in the Process of Learning to Read. *Asia-Pacific Journal of Teacher Education, 33*, 65-75.

Fowler, A. E., & Swanson, B. (2005). Relationships of Naming Skills to Reading, Memory, and Receptive Vocabulary: Evidence for Imprecise Phonological Representations of Words by Poor Readers. *Annals of Dyslexia, 4*(2), 247-280.

Frith, U. (1997). Brain, mind and behaviour in dyslexia. In C. Hulme and M. Snowling (Eds.), *Dyslexia: Biology, cognition and intervention*. London: Whurr.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical Analysis. *Scientific Studies of Reading, 5*, 239-256.

Garside, A. H., & Wilkins, A. M. (2002). *Basic Facts About Dyslexia: What Every Layperson Ought to Know*, 10-11. Baltimore: International Dyslexia Association Pamphlet series.

Goforth, C. (2015). *Using and Interpreting Cronbach's Alpha*. http://data.library.virginia.edu/ using-and-interpreting-cronbachs-alpha/

Ho, C. S. H. (2010). Understanding reading disability in Chinese: From basic research to intervention. In M. H. Bond (Ed.), *Handbook of Chinese psychology* (2nd Ed., pp. 109-121). New York: Oxford University Press.

Jenkins, J. R., Johnson, E., & Hileman, J. (2004). When is reading also writing: Sources of individual differences on the new reading performance assessments. *Scientific Studies of Reading, 8*, 125-151.

Joshi, N. A., & Vanaja, C. S. (2016). *Checklist to Screen Children with Reading Difficulty (CSRD) for Classroom Teachers' Language in India*. www.languageinindia.com 16:1.

Lange, S. M., & Thompson, B. (2006). Early Identification and Interventions for Children at Risk for Learning Disabilities. *International Journal of Special Education, 21*(3), 108-119.

Layes, S., Lalonde, R., & Rebaï, M. (2014). Reading Speed and Phonological Awareness Deficits among Arabic-Speaking Children with Dyslexia. *Dyslexia, 21*(1), 80-95.

Layes, S., Lalonde, R., Mecheri, S., & Rebaï, M. (2015). Phonological and Cognitive Reading Related Skills as Predictors of Word Reading and Reading Comprehension among Arabic Dyslexic Children. *Psychology*, *6*, 20-38.

Layes, S., Lalonde, R., & Rebaï, M. (2016). Study on Morphological Awareness and Rapid Automatized Naming Through Word Reading and Comprehension in Normal and Disabled Reading Arabic-Speaking Children. *Reading & Writing Quarterly, 33*(2), 123-140.

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of Dyslexia. *Annals of Dyslexia, 53*(1), 1-14.

Mash, E. J., & Terdal, L. G. (Eds.). (1997). *Assessment of childhood disorders* (3rd Ed.). New York: Guilford.

Mather, N., & Wendling, B. J. (2012). *Essentials of dyslexia assessment and intervention.* Hoboken, NJ: John Wiley & Sons.

National Reading Panel. (2000). *Teaching Children to Read: Evidence-Based Assessment of the Scientific Research Literature on Reading and its Implications for Reading Instruction*. National Institute of Child Health and Human Development; Bethesda, MD.

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of Single-Word Reading and Text Comprehension: Evidence from Component Skills. *Language and Cognitive Processes*, *18*, 443-468.

Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading*, *11*, 357-383.

Pugh, K. R., Sandak, R., Frost, S. J., Moore, D., Rueck J. G., & Mencl, W. E. (2013). Neurobiological studies of skilled and impaired reading: A work in progress. In G. D. Rosen (Ed.)., *The Dyslexic Brain: New Pathways in Neuroscience Discovery* (pp. 21-47). New York: Psychology Press.

Rief, S. F., & Stern, J. M. (2010). The *dyslexia checklist: A practical reference for parents and teachers*. San Francisco: John Wiley.

Salvesen, K. A., & Undheim, J. O. (1994). Screening for Learning Disabilities. *Journal of Learning Disabilities*, *27*, 60-66.

Schulte-Körne, G., Deimel, W., & Remschmidt, H. (1997). Can Self-Report Data on Deficits in Reading and Spelling Predict Spelling Disability as Defined by Psychometric Tests? *Reading and Writing*, *9*, 55-63.

Shankweiler, D., Katz, L., Stuebing, K., Fletcher, J. M., Brady, S., Fowler, A., et al. (1999). Comprehension and Decoding: Patterns of Association in Children with Reading Difficulties. *Scientific Studies in Reading*, *3*, 69-94.

Shelton, T. L., & Barkely, R. A. (1994). Critical Issues in the Assessment of Attention Deficit Disorders in Children. *Topics in Language Disorder. Special Issue: ADD and its relationship to spoken and written language, 14*, 26-41.

Snow, C. E. (2002). *Reading for understanding: Toward a research and development program in reading comprehension*. Pittsburgh: RAND.

Speece, D. L., Ritchey, K. D., Silverman, R., Schatsneider, C., Walker, C. Y., & Andrusik, K. N. (2010). Identifying Children in Middle Childhood Who are at Risk for Reading Problems. *School Psychology Review*, *39*(2), 258-276.

Wadlington, E. M., & Wadlington, P. L. (2005). What Educators Really Believe about Dyslexia? *Reading Improvement, 42*(1), 16-33.

Wagner, W. G. (2003). *Counselling, psychology, and children: A multidimensional approach to intervention.* Upper Saddle River, NJ: Merrill.

Willcutt, E. G., Petrill, S. A., Wu, S., Boada, R., DeFries, J. C., Olson, R. K., & Pennington, B. F. (2013). Comorbidity Between Reading Disability and Math Disability Concurrent Psychopathology, Functional Impairment, and Neuropsychological Functioning. *Journal of Learning Disabilities*, *46*(6), 500-516.

Willcutt, E. G., Boada, R., Riddle, M. W., Chhabildas, N., DeFries, J. C., & Pennington, B. F. (2011). Colorado Learning Difficulties Questionnaire: validation of a Parent-Report Screening Measure. *Psychological Assessment*, *23*(3), 778-791.

Wolff, U., & Lundberg, I. (2003). A Technique for Group Screening of Dyslexia among Adults. *Annals of Dyslexia*, *53*(1), 324-339.

Wong, A. M. Y., Ho, C. S-H., Au, T. K-F., McBride, C., Ng, A. K-H., Yip, L. P-W., Lam, C. C-C. (2017). Reading Comprehension, Working Memory and Higher-Level Language Skills in Children with SLI and/or Dyslexia. *Reading and Writing, 30*, 337-361. Doi: 10.1007/s11145-016-9678-0.

Received July 3, 2018
Revision October 1, 2018
Accepted October 19, 2018