Journal of Research in Reading
UKLA
The United Kingdom Literacy Association

Do reading comprehension assessment tests result in the same reading profile? A study of Spanish primary school children

Journal:	*Journal of Research in Reading*
Manuscript ID	JRIR-2018-12-0142.R3
Manuscript Type:	Original Article
Keywords:	Assessment, standardized reading comprehension tests, poor comprehenders, primary school children, reading comprehension difficulties

Abstract

Background: Detecting reading comprehension difficulties is challenging because many factors are involved in comprehension ability. Various reading comprehension tests can be used to detect difficulties but often do not yield the same results.

Method: Our aim was to analyse the agreement between three commonly used standardised reading comprehension tests (ECOMPLEC, ACL and PROLEC-R) in the detection of reading comprehension difficulties in Spanish. A total of 139 children (72 fifth graders and 67 sixth graders) at the same public-sector school participated in this study. The three reading comprehension tests were administered, together with word and nonword reading, vocabulary and nonverbal intelligence measures.

Results: Modest intercorrelations among the tests were found. The consistency of classification for each reading profile across the three reading comprehension tests was low. The results show different reading comprehension profiles depending on the test used.

Conclusions: It is important to use more than one instrument to diagnose reading comprehension difficulties, due to the complexity involved. Furthermore, knowledge of the characteristics of each reading comprehension test is essential to the choice of test. The educational implications of children being wrongly diagnosed are discussed.

Keywords: reading comprehension difficulties; assessment; standardised reading comprehension tests; poor comprehenders; primary school children

What is already known about this topic

• Accurate detection of reading comprehension difficulties in primary school children is essential to ensure early and effective intervention.

• Assessing reading comprehension is challenging because many cognitive processes are involved in this complex ability.

• Widely used English reading comprehension tests are known to produce varying results.

What this paper adds

• This is the first study to analyse the difference in results obtained from three widely used Spanish-language standardised reading comprehension tests (ECOMPLEC, PROLEC-R and ACL).

• Moderate intercorrelations were observed between the three reading comprehension tests.

• Important differences were found in reading comprehension classifications for the three reading comprehension tests.

Implications for theory, policy or practice

• It is important to use more than one measure of reading comprehension in order to diagnose reading comprehension difficulties in Spanish.

• It is essential to be aware of what cognitive processes are being assessed by a particular reading test.

• The manuals for reading comprehension tests should indicate the processes related to comprehension that are being assessed, to help interpret the data.

Introduction

Comprehension is the ultimate goal of reading and many studies have reported a significant positive correlation between reading comprehension and academic achievement (Elosúa et al., 2012; García-Madruga, Vila, Gómez-Veiga, Duque & Elosúa, 2014; Meneguetti, Carretti & De Beni, 2006; O'Reilly & McNamara, 2007). This skill is also relevant for personal and social development (Gray, 2017; Kamil, 2003). Nevertheless, and despite its importance, many children struggle to understand what they are reading. As pointed out by Catts, Compton, Tomblin and Bridges (2012), a significant proportion of children (around 16% in grades 2 to 10) can be categorised as poor readers. Similarly, Clarke, Snowling, Truelove and Hulme (2010) reported that around 10% of primary school children have reading comprehension difficulties. Regarding Spanish primary pupils, the Progress in International Reading Literacy Study (PIRSL) reports for 2011 (Mullis, Martin, Foy & Drucker, 2012) and 2016 (Mullis, Martin, Foy & Hooper, 2017) indicate that reading comprehension levels are below the OECD average (OECD, 2017). Furthermore, García, Jiménez, González and Jiménez-Suárez (2013) report that approximately 20% of the Spanish population have reading comprehension difficulties.

Reading comprehension assessment: factors involved

Accurate detection of reading comprehension difficulties in children is essential both to ensure early and effective intervention and to facilitate empirical research (Cain & Oakhill, 2006a). However, while the assessment of decoding skills is well established (typically measuring accuracy in terms of words/pseudowords read per minute), the assessment of reading comprehension is more complicated and challenging because many cognitive processes are involved in this complex ability (Cain, 2016; Francis, Fletcher, Catts & Tomblin, 2005). It seems

difficult to grasp in one test all the cognitive processes involved in reading comprehension. Moreover, other factors must also be taken into account; for example, reflecting the complexity of reading comprehension skills, Francis et al. (2005) showed that the relation between decoding and oral languages with reading comprehension skills varies with age, and so this variable must be taken into consideration (Catts, Hogan & Adolf, 2005; Keenan et al., 2008).

Reading, as proposed by the simple view of reading, is a product of decoding and of linguistic comprehension (see Gough & Tunmer, 1986). Taking into account these two main factors, reading comprehension difficulties can be seen as inadequacies in decoding and in comprehension (Hoover & Gough, 1990). Children with poor reading comprehension are usually identified using percentiles or standardised score cutoffs for achievement in word reading and reading comprehension (Adlof, Catts & Lee, 2010). To this end, a large number of standardised reading comprehension tests have been proposed, in various languages. Many of these are used interchangeably because they are presented as measures of the same construct of reading comprehension. However, given that reading comprehension consists of multiple cognitive processes, some questions arise. Are these measures equally effective in detecting poor comprehenders? Do they all measure the same comprehension skills component? Do they offer the same reading comprehension profile?

Reading comprehension test differences

Research shows that a different reading profile may be obtained depending on the reading comprehension tests used. One of the first studies that compared different reading comprehension tests was carried out by Nation and Snowling (1997), who studied 189 British children aged 7-10 years and compared two tests: the Neale Analysis of Reading Ability (NARA; Neale, 1989) and

the Suffolk Reading Scale (SRS; Hagley, 1987). The authors concluded not only that different reading comprehension tests assessed different aspects of the reading process, but that listening comprehension accounted for additional variance only in the NARA, and that performance in both tests was influenced by decoding skills. In the USA, in a study of 97 first-to-tenth graders, Cutting and Scarborough (2006) analysed three widely used tests — the Wechsler Individual Achievement Tests (WIAT; Wechsler, 1992), the Gates–MacGinitie Reading Test (GMRT; MacGinitie, MacGinitie, Maria & Dreyer, 2000), and the Gray Oral Reading Test (GORT; Wiederholt & Bryant, 1992) — and reported inconsistencies between the tests in terms of identifying which children had comprehension difficulties. Other studies have corroborated that correlations between scores for reading comprehension assessments are surprisingly low and that different reading comprehension tests are inconsistent in their diagnoses (Colenbrander, Nickels & Kohnen, 2017; Collins, Lindström & Compton, 2018; Keenan, Betjemann & Olson, 2008; Keenan, et al., 2014; Keenan & Meenan, 2014). For example, Keenan and Meenan (2014) assessed 995 children (mean age 11.17 years) using four standardised reading comprehension tests: the GORT-3 (Wiederholt & Bryant, 1992), the Qualitative Reading Inventory-3 (QRI-3; Leslie & Caldwell, 2001), the Woodcock-Johnson Passage Comprehension-3 (WJPC-3; Woodcock, McGrew & Mather, 2001) and the Peabody Individual Achievement Test (PIAT-3; Dunn & Markwardt, 1970). The median correlation between these tests was only .54.

Differing results in reading comprehension tests have also been observed in studies of Spanish-language reading comprehension. Thus, López-Escribano, Elosúa, Goméz-Veiga & García-Madruga (2013) used an unpublished test called Diagnostic Assessment Reading Comprehension (DARC; Francis et al., 2006) and the PROLEC-R revised battery for the evaluation of reading processes (Cuetos, Rodríguez, Ruano & Arribas, 2007) to analyse the

contribution of language and cognitive skills to reading comprehension in third grade primary school children. These authors measured a low correlation (.37) between the two tests. Moreover, the variables that predicted reading comprehension differed between the tests. Pseudo-word reading time was important for PROLEC-R, whereas verbal working memory and spelling skills predicted DARC scores. In another study, Elosúa et al. (2012) applied DARC (Francis et al., 2006) and PROLEC-R (Cuetos et al., 2007) to 3rd and 6th grade children, and obtained a low correlation (.36) between the two measures. PROLEC-R produced lower correlations than DARC with the other measures in the study (decoding and academic achievement). These authors concluded that the two tests were different in terms of structure and response format which could be the cause of the low correlation found between the results of the tests.

To explain these differences, some researchers indicate that the different reading comprehension tests do not assess the same array of cognitive processes (Fletcher, 2006). Others state that the reading comprehension tests may differ in factors such as the presentation structure (e.g., whether the text is available while answering the questions, whether the text can be consulted, text length, question type, etc) and in the way they are administered (e.g., multiple choice, open questions, short answers, retell, timed answers, etc; for a review, see Collins et al., 2018). These factors may influence the reading comprehension scores obtained. For instance, in the PIAT test (Dunn & Markwardt, 1970), children read a sentence, then have to select a single picture among four that best represents the sentence's meaning; thus, correct word decoding is essential to choosing the correct answer. This is also the case with the WJPC (Woodcock et al., 2001) — a cloze format test — whereby the right answer also requires the correct decoding of a single word. However, other formats that use longer texts are less dependent on decoding skills. In this case, a child with decoding problems might nevertheless achieve the correct answer if the context

compensated for the decoding problems (Miller et al., 2008) (for a review see Keenan et al., 2008; Keenan, et al., 2014; Keenan & Meenan, 2014).

Different scores have even been found in training studies, depending on the reading comprehension test used. Clarke et al. (2010) conducted a study aimed at improving reading comprehension in children aged 8-9 years using two reading comprehension tests: NARA II (Neale, 1997) and WIAT (Wechsler, 2005). Although all the children received the same training, only one of these tests reflected a significant improvement in reading comprehension. In Spain, too, intervention studies have found differences in reading comprehension depending on the test applied. For example, Calet, Guitiérrez-Palma & Defior (2017) found that fluency training had an impact at the sentence comprehension level but not at the passage comprehension level.

This discrepancy in reading comprehension test scores is worrying, as children may be wrongly diagnosed, with potentially damaging academic and personal consequences (Cain & Oakhill, 2006b; McLaughlin, Speirs & Shenassa, 2012; Ricketts, Sperring & Nation, 2014). Apart from the question of discrepancies in reading comprehension test results, diagnosis itself is complicated by the fact that different criteria are used across research and clinical settings (Clarke, Henderson & Truelove, 2010) and the diagnosis of poor comprehenders is sometimes controversial. Some researchers consider only reading comprehension scores and do not take decoding skills into account, whereas others require children to have age-appropriate decoding scores or require a set discrepancy between reading comprehension and decoding scores (Keenan, et al., 2014).

The present study

In short, studies of widely-used English reading comprehension tests indicate that they produce different results, possibly because they do not reflect the same cognitive processes. To our knowledge, little research has been conducted in a Spanish-language context to analyse the use of different reading comprehension tests in primary school children. In the view that it is important to determine the strengths and weakness of different measures of reading comprehension, the present study has three main aims:

(1) to determine the level of agreement between three widely used Spanish-language standardised reading comprehension tests (ECOMPLEC, PROLEC-R and ACL) in schools and in research.

(2) to determine the extent to which reading comprehension scores are predicted by oral skills (vocabulary) and/or decoding scores.

(3) to explore whether the profile of reading difficulties differs across tests and to determine whether the different instruments are equally effective in detecting reading difficulties.

Method

Participants

The sample comprised 139 children from a public-sector school in the south of Spain with Spanish as their first language: 72 were fifth graders (mean age = 124.19 months, *SD* = 3.4 months; 29 boys and 43 girls) and 67 were sixth graders (mean age = 136.58 months, *SD* = 3.7 months; 29 boys and 38 girls). The children in the sample attended a mainstream school, and presented a continuum of reading levels. None had any known cognitive impairment, visual or motor disorders, and all were native Spanish speakers.

Measures

Three standardised Spanish-language reading comprehension tests were applied. In all cases, the children had unlimited time to read the texts and answer the questions. They were also assessed for intelligence (nonverbal and verbal skills) and decoding skills.

Reading comprehension skills

ECOMPLEC. Reading Comprehension Assessment Test (León, Escudero & Olmos, 2012). This test consists of three texts (narrative, expository and discontinuous) with multiple-choice questions (three possible answers). The narrative text had 542 words and 24 questions, the expository passage had 348 words with 23 questions and the discontinuous text presented a web format, with 170 words and 26 questions. The children were told to read the texts in silence and then to answer 36 inferential questions, 31 literal questions and 6 metacognitive questions. Each correct answer scored one point and the maximum score possible was 67 points (the metacognitive questions were not scored). The manual reports a Cronbach's alpha reliability coefficient of .89.

Reading comprehension test: PROLEC-R. A subtest of the revised battery for the evaluation of reading processes (Cuetos et al., 2007) was used as another measure to assess reading comprehension. This test consists of two narratives (with 94 and 131 words, respectively) and two expository texts (with 75 and 138 words). The children were told to read the texts in silence and then to answer 16 open-ended inferential questions (four per text). A written response was required. Each correct answer scored one point. For an answer to be considered correct, only the semantic content was taken into account, not the spelling or other characteristics unrelated to the comprehension of the text. The Cronbach alpha reliability coefficient obtained was .72.

Assessment of reading comprehension: ACL test (Català, Català, Molina & Monclús, 2001). This test is divided into six reading levels, each corresponding to a primary education level

in Spain (first grade to sixth grade). The text includes narrative, expository, data interpretation and graphical interpretation texts. For this study, we used levels 5 and 6. In both levels, the children have to silently read three expository, one data interpretation, two graphical interpretation and one narrative text (ten in total), then answer 35 (level 5) or 36 (level 6) multiple-choice questions, with five possible answers and a score of one point for the correct answer. For every passage, there were three, four or five questions. The texts for the level 5 test had 25-164 words, and those for level 6, 41-170 words. The manual reports a Cronbach's alpha reliability coefficient of .82 for level 5 and .76 for level 6.

General cognitive ability

Kaufman Brief Intelligence Test (K-BIT). The Spanish version of K-BIT (Kaufman & Kaufman, 1997) is an individually-administered measure of verbal and nonverbal abilities in individuals aged 4-90 years. It is composed of two subtests: Vocabulary (verbal intelligence) and Matrices (nonverbal intelligence). In turn, the verbal scale is composed of two subtests, Expressive Vocabulary and Definitions, while the nonverbal scale is composed of one subtest, Matrices.

Expressive Vocabulary requires participants to name a series of pictured objects. The test starts with high frequency words (for example: bed, fork, stairs…) and continues with items presenting lower frequency (for example: harpoon, hexagon, anvil…). *Definitions* (administered only to individuals aged 8 years or older) requires the identification of words that best fit two clues; a brief verbal description and a partial spelling of the word. For example, children must answer the word *antiguo* (ancient) when we present the partial spelling "_NT_G_ _" and the description "something very old".

Matrices is a multiple-choice task that, in general, requires the recognition of relationships among visual stimuli. This subtest includes several item types, all of which involve understanding relationships among concrete and abstract visual stimuli. The easiest items require selecting the picture that best matches visual stimuli, while more complex items require solving visual analogies utilising abstract stimuli (more details in Wang & Kaufman, 1993).

Each correct answer is scored with 1 point and the maximum score is 130 points. The manual reports a Cronbach's alpha reliability coefficient of .94 (vocabulary subtest) and .86 (matrices subtest).

Decoding skills

Word reading. A subtest of the PROLEC-R Spanish standardised battery (Cuetos et al., 2007) was used to measure word reading ability. In this subtest, individuals are told to read a list of 40 words aloud as quickly and as accurately as they can. These words vary in frequency, length and syllabic structure (CCV, CVV, CVC, CCVC, CVVC and VC, where C = consonant and V = vowel). The accuracy score is calculated as one point awarded for each word read correctly. The reading rate, i.e., words correctly read aloud per minute, is scored by dividing the accuracy score by the total reading time (in seconds) and multiplied by 60. Cronbach's alpha coefficient indicated by the manual is .74.

Nonword reading. This subtest, also included in the PROLEC-R Spanish standardised battery (Cuetos et al., 2007), requires children to read 40 nonwords as quickly and as accurately as possible. The accuracy score is calculated as one point awarded for each nonword read correctly. The nonword reading rate, i.e., nonwords correctly read aloud per minute, was scored by dividing the accuracy score by the total reading time (in seconds) and multiplied by 60, yielding the number

of correct nonwords per minute. Cronbach's alpha coefficient indicated by the manual is .68. This coefficient can be considered moderate. Therefore, we calculated the alpha for the present study sample, obtaining a result of .72.

Academic achievement

To assess academic achievement, the students' final semester examination marks (range 0-10) were obtained from the school records. The teachers provided the final semester examination scores in the basic subjects (mathematics, language, science, arts, social science and foreign language), from which the mean was calculated.

Procedure

The assessment took place in three sessions (two group sessions and one individual session) lasting no longer than 60 minutes each. The tests were counterbalanced in order to control for possible effects due to the order of presentation of the tests. Evaluations were performed by trained examiners. In accordance with the rules of the Ethical Human Research Committee of the University of Granada (Spain), participation forms were received from the participating school and data were only collected from children whose parents had returned signed consent forms.

Data analytic methods

Descriptive statistics were used to describe our sample. Correlations were calculated for the analysis of the relations among the reading comprehension tests and to interpret the scores for decoding, vocabulary and academic achievement. Hierarchical linear regression analyses were then performed for each of the three reading comprehension tests to determine how much of the variance in test scores was accounted for by differences in decoding skills and vocabulary. The

prevalence values for reading profiles obtained by each of the reading comprehension tests, using decoding measures, were also calculated to examine whether the reading difficulty profiles were similar in each test. Finally, the consistency of classification among the three reading comprehension tests was analysed to determine whether the different instruments were equally effective in detecting reading difficulties.

Results

Table 1 summarises the descriptive statistics obtained, showing that the overall performance of the participants was slightly above average. The distribution plots for each measure were normal. The measure of nonverbal intelligence measure was assessed purely for descriptive purposes, and so is not included in further analysis.

Insert Table 1 about here

Table 2 shows the correlation matrix for the three reading comprehension tests, the word and nonword reading rates and the scores for vocabulary and academic achievement. These results reflect a significant positive correlation among the three reading comprehension measures. However, the correlations between PROLEC-R and the other two reading comprehension tests were weak, although those between ECOMPLEC and ACL were stronger. Overall, the correlations among the three reading comprehension tests were modest, given that they all seek to measure the same construct.

The word reading rate was strongly correlated with the nonword reading rate ($r = .65$, $p < .001$). Both of these rates were significantly correlated with the three reading comprehension tests. The results of the vocabulary test also presented significant positive correlations with the three

reading comprehension tests. This was only to be expected, in view of the growing body of evidence linking vocabulary to reading comprehension abilities (e.g., Baumann, 2009; Perfetti, 2007; Perfetti & Hart, 2001). There were moderate positive correlations between academic achievement and ECOMPLEC ($r = .40, p < .001$) and between academic achievement and ACL ($r = .41, p < .001$).

Insert Table 2 about here

Hierarchical linear regression analyses were performed to analyse how much of the variance in reading comprehension test scores was accounted for by decoding skills and vocabulary. As both word and nonword reading rates were strongly correlated, a composite score was calculated in order to obtain a single measure for decoding skills. This composite score was the result of the average of z-scores of word reading rate and nonword reading rate measures.

Insert Table 3 about here

Age was initially entered as a control variable, followed by vocabulary and decoding skills. As can be seen in Table 3, the results show that after discounting the effects of age, vocabulary skills accounted for 23% of the variance in both the ECOMPLEC and the ACL scores, but only 10% of the variance in the PROLEC-R scores. Decoding skills, on the other hand, accounted for 26% of the variance in the ECOMPLEC scores, but for no additional variance in the PROLEC-R

or ACL scores. Therefore, it seems that the ECOMPLEC scores are more dependent on the individual's decoding skills.

The different reading comprehension tests and decoding measures were also used to obtain the following order of prevalence for the reading profiles (in relation to age): successful readers, poor readers, poor comprehenders and poor decoders (Table 4). Successful readers had average or above average decoding and comprehension scores, whereas poor readers had below average decoding and comprehension scores. Poor decoders had below average decoding skills but average comprehension skills, whereas poor comprehenders had above average decoding skills but below average comprehension skills. Although the notion of poor comprehender is controversial, we emphasise that this category reflected children who performed poorly in comprehension but had good decoding skills. To calculate the prevalence of each reading profile, we used a cutoff criterion of 1.5 standard deviations below the mean, which is consistent with the DSM-5 standard (American Psychiatric Association, 2013).

Insert Table 4 about here

As shown in Table 4, the prevalence of each profile varied depending on the reading comprehension test used, and no consistent prevalence rate was obtained across the three tests. Although most children in our sample were successful readers, more were diagnosed as poor comprehenders by PROLEC-R than by ECOMPLEC or ACL. Using the word or nonword reading rates as decoding measures made little difference to the prevalence figures.

The consistency of classification across the three reading comprehension tests was also calculated. The consistency of diagnosis— i.e., the proportion of children with the same reading

profile according to the three reading comprehension tests (ECOMPLEC, PROLEC-R and ACL) — is summarised in Table 5. In this table, the consistency of classification for each reading profile across the three reading comprehension tests can also be found.

Insert Table 5 about here

As can be seen in Table 5, overall consistency between the three tests varied from 79% to 81%, depending on which decoding assessment was used, with somewhat higher consistency when the nonword reading rate was used as a decoding measure.

The consistency between the three tests in the reading profile of poor comprehenders varied from 4% to 8%, depending on which decoding assessment was used, with somewhat higher consistency when the word reading rate was used as a decoding measure. Consistency between the three tests in the reading profile of poor readers varied from 0% to 20%, while for poor decoders, it varied from 43% to 62%. Finally, consistency between the three tests in the reading profile of successful readers varied from 82% to 81%. Overall consistency was somewhat higher when the nonword reading rate was used, with the exception of the profile of poor comprehenders.

Discussion

The aim of this study was to determine the level of agreement between three standardised reading comprehension tests (ECOMPLEC, PROLEC-R and ACL), which are commonly used to detect reading comprehension difficulties in Spanish, using a sample of primary school children. To our knowledge, this is the first study undertaken to analyse the results obtained by various

standardized Spanish-language tests of reading comprehension; the majority of previous studies in this field have compared tests of English-language reading comprehension.

An important finding of this study is that only low to moderate intercorrelations were observed among the three reading comprehension tests, corroborating previous studies in this field (Colenbrander et al., 2017; Keenan et al., 2008; Keenan, et al., 2014; Keenan & Meenan, 2014). This finding suggests that the tests examined do not measure the same construct, and that they may reflect different cognitive processes. The lowest correlation was found between PROLEC-R and the other two tests (r =. 30-.33). This test was also poorly correlated with academic achievement. This result is in line with previous studies in Spain, which have reported low correlations between PROLEC-R and DARC (Elosúa, et al., 2012; López-Escribano et al., 2013). Moreover, Elosúa et al. (2012) found that the correlation with academic achievement was lower with PROLEC-R than with DARC. There are several reasons why a higher correlation was obtained between the ECOMPLEC and ACL scores: both are based on multiple-choice questions, whereas PROLEC-R includes open-ended questions. While multiple-choice tests are popular in research and are suitable for group administration, the processing demands are higher because a variety of options must be compared (Cain & Oakhill, 2006a); this was especially the case of ACL, with five response alternatives. In contrast, open answers — like PROLEC-R and English-language tests such as NARA (Neale, 1999) — require output demands; these verbal responses require additional cognitive effort in terms of structuring and formulating a response, which may confound the assessment of comprehension if a poor comprehender has expressive language difficulties.

Another difference between the tests is that of the question type, which is likely to affect consistency between tests (Bowyer-Crane & Snowling, 2005). Thus, while both ECOMPLEC and ACL have literal, inferential, metacognitive questions, etc, PROLEC-R has only inferential

questions. On the other hand, presenting different types of question may enhance the precision with which the complex process of comprehension is measured. Finally, the tests differ in the number of questions presented, with PROLEC-R having far fewer questions (16) than ECOMPLEC (67) and ACL (35-36). Therefore, ECOMPLEC and ACL offer a more precise estimate of comprehension given the tests include more questions. These considerations lead us to conclude that the reading comprehension score depends to some extent on the response format adopted, as observed previously by Francis et al. (2005).

The hypothesis that similar cognitive processes are measured by ECOMPLEC and ACL is supported by the fact that the amount of variance accounted for by vocabulary was the same in each case (.23), and greater than for PROLEC-R (.10). The vocabulary scores for all the tests were significant, which is not surprising given the well-known relationship between vocabulary and reading comprehension skills (Cain, 2016; Cain & Oakhill, 2006a; Cain & Oakhill, 2006b).

Another interesting finding is that the ECOMPLEC comprehension scores were more dependent on decoding skills than were those obtained by PROLEC-R and ACL. Therefore, when tested with ECOMPLEC, children with better word reading rates obtained higher comprehension scores. Since the ECOMPLEC passages were longer than in the PROLEC-R and ACL tests, it was necessary to read rapidly and accurately in order to free up resources for comprehension. This conclusion is in line with the theory of automaticity (LaBerge & Samuels, 1974), according to which word reading fluency is a relevant skill for predicting reading comprehension (Adlof, Catts & Little, 2006; LaBerge & Samuels, 1974; Perfetti, 1985). It would be interesting to explore this relationship in populations of older and younger children, as previous studies have found that correlations between decoding and reading comprehension are higher in early grades than in later ones (Francis et al., 2005).

The ECOMPLEC passages also appeared to be more linguistically challenging than those in the PROLEC-R and ACL tests, requiring good fluency skills. Previous studies of readers of English have reported that tests consisting of sentences tend to be more influenced by decoding than tests with long passages. Thus, Keenan et al. (2008) reported that the PIAT test (Dunn & Markwardt, 1970) — consisting of single sentences — and the WJPC test (Woodcock et al., 2001) — based on a cloze format — were heavily influenced by decoding skills compared with those tests with medium and long passages (GORT and QRI tests). These authors argued that in longer passages the context may help the child to correctly decode challenging words. However, by looking at PIAT and WJPC test items in close detail, it can be seen that word decoding is essential for choosing the correct answer in these two tests, while this skill is less relevant in GORT and QRI. Another factor to consider is a developmental issue. In the study by Keenan et al. (2008) the sample included children across a broad age range (from 8 to 18 years). These authors found that decoding skill was less important in PIAT and WJPC for more skilled and older children. Besides, they also explained that decoding plays a small role in the GORT test because the examiner (not the child) reads the test questions.

To explain the contrasting findings with previous studies English speaking children it is also necessary to take into account the language characteristics. In the present study, in contrast with Keenan et al. (2008), we included speed in the decoding measure. This is important in a transparent language, such as Spanish, in which children typically achieve a reading accuracy of 95% of words (Seymour, Aro & Erskine, 2003). In fact, studies in transparent systems often rely on reading speed as indicator of individual differences (Cuetos, & Suárez-Coalla, 2009). In future Spanish studies it would be interesting to analyse reading comprehension tests using long passages, medium passages, with sentences and taking into account more reading levels.

In agreement with studies focused on English-language reading comprehension (Colenbrander et al., 2017; Collins et al., 2018; Keenan et al., 2008; Keenan, et al., 2014; Keenan & Meenan, 2014), we found important differences in classifications obtained by the three reading comprehension tests, despite their similar formats. One explanation for this may be that children's comprehension scores vary because reading comprehension is a complex ability involving many cognitive processes. A multitude of language skills, such as inference, comprehension monitoring and knowledge and use of story structure (Oakhill & Cain, 2012) are required and may influence reading comprehension performance in any given test. The three tests used in our study may reflect different language skills. For example, a child may require more knowledge and use of a story structuring ability to achieve success in ACL, with a variety of structured texts, than in the ECOMPLEC or PROLEC-R tests. Measures of inferential capacity, comprehension monitoring and narrative production could be examined in future work in this field to determine which language factors related to comprehension make the greatest contribution to differences in reading profiles.

Regarding consistency, previous reviews highlighted that consistency rates from prior research (e.g., 35%-65%; Keenan et al., 2008; Keenan & Meenan, 2014) seem lower than the estimates in the present study (79%-81%). Yet, the Colenbrander et al., (2017) consistency rates of 66-85% seem comparable to the present study, possibly because the two tests that they used (the NARA test and the York Assessment of Reading for Comprehension), are relatively similar in format. Although our overall consistency rate seems high, the data, when examined in detail concerning the consistency for each reading profile, reveal that the proportion of children in the poor comprehender category was actually extremely low (4-8%). This is worrying, given that according to Nation, Snowling & Clarke (2005) poor comprehenders are relatively common but

often go unnoticed due to the difficulties in detecting them. Poor comprehenders read aloud accurately, but their difficulties in comprehension may have serious educational consequences (Hulme & Snowling, 2011).

Our findings support the idea that several measures of reading comprehension are necessary for the proper diagnosis of reading comprehension difficulties in Spanish (Bowyer-Crane & Snowling, 2005; Cain & Oakhill, 2006a; Colenbrander et al., 2017; Keenan & Meenan, 2014), as a single measure may not capture the complexity of reading comprehension abilities (Catts & Kamhi, 2017). As shown in our study, ECOMPLEC, PROLEC-R and ACL vary in terms of the assessment format used and in the components of comprehension skills that they measure. In deciding on the appropriateness of a given test, therefore, it is important to analyse and fully understand its characteristics.

The findings of our study have important implications both for clinical research and for educational assessment. As Cain and Oakhill (2006a) point out, no assessment tool is perfect. Nonetheless, given that tests differ in many factors, including the length of texts, the type of questions used, the cognitive processes assessed and the assessment format, it is important to be familiar with the characteristics of each reading comprehension test, in order to select the right test for our needs and to interpret the test scores. If this is not done, some children may continue through school with unrecognised difficulties, giving rise to serious academic and personal consequences.

We realise that limited time is available for testing, whether in research or in school settings. Nevertheless, we recommend that at least two reading comprehension tests be used or, at least, that the user should know which cognitive processes are being assessed by a particular test. It would be useful for reading comprehension tests to include information on the cognitive

processes and skills evaluated so that researchers and teachers can properly interpret the data. Manuals for reading comprehension tests usually include important statistical information such as external validity, internal validity and reliability. However, our findings show that these manuals should also indicate the processes related to comprehension, such as vocabulary or decoding, that are tested by each measure of reading comprehension.

The present study has some limitations that should be noted. The research was focused on the children in two grades at a single public-sector school. In future research, it would be useful to include different kind of schools. On the other hand, three different classes were assessed for each grade within the school, and therefore we believe our sample is representative.

The inclusion of only fifth and sixth-grade children is another limitation. For the conclusions obtained to be more generalisable, a broader-based population should be analysed. In future research, it would be interesting to consider younger, primary-level children, and also to conduct longitudinal studies in order to explore the predictive abilities of the different Spanish tests analysed in this paper. Furthermore, given that many standardised reading comprehension tests are available in Spanish, another research line could be to analyse tests with different formats.

References

Adlof, S. M., Catts, H. W. & Lee, J. (2010). Kindergarten predictors of second versus eighth grade reading comprehension impairments. *Journal of Learning Disabilities, 43*, 332–345. https://doi.org/10.1177/0022219410369067.

Adlof, S., Catts, H W. & Little, T. (2006). Should the simple view of reading include a fluency component? *Reading & Writing, 19*(9), 933-958. https://doi.org/10.1007/s11145-006-9024-z.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*. (5th ed). Arlington, VA: American Psychiatric Publishing.

Baumann, J. F. (2009). Vocabulary and reading comprehension: The nexus of meaning. In: S. E. Israel and G.G. Duffy (editors). *Handbook of Research on Reading Comprehension*. (pp. 323–346). New York, NY: Routledge.

Bowyer-Crane, C. & Snowling, M.J. (2005). Assessing children's inference generation: What do tests of reading comprehension measure? *British Journal of Educational Psychology, 75*, 189–201. https://doi.org/10.1348/000709904X22674.

Cain, K. (2016). Reading comprehension development and difficulties: an overview. *Perspectives on Language and Literacy, 42*(2), 9-16.

Cain, K. & Oakhill, J. (2006a). Assessment matters: Issues in the measurement of reading comprehension. *British Journal of Educational Psychology, 76*, 697–708. https://doi.org/10.1348/000709905X69807.

Cain, K. & Oakhill, J. (2006b). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology, 76*, 683–696. https://doi.org/10.1348/000709905X67610.

Calet, N., Gutiérrez-Palma, N. & Defior, S. (2017). Effects of fluency training on reading competence in primary school children: The role of prosody. *Learning and Instruction, 52*, 59-68. https://doi.org/10.1016/j.learninstruc.2017.04.006.

Català, G., Català, M., Molina, E. & Monclús, R. (2001). *Evaluación de la comprensión lectora: Pruebas ACL* [Assessment of reading comprehension: ACL tests]. Barcelona (Spain): Editorial Graó.

Catts, H. W., Compton, D. L., Tomblin, J. B. & Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology, 104*, 166–181. http://dx.doi.org/10.1037/a0025323.

Catts, H.W., Hogan, T. P. & Adolf, S. M. (2005). Developmental changes in reading and reading disabilities. In H.W. Catts & A. G. Kamhi (Eds.), *The connections between language and reading disabilities*. Mahwah, NJ: Erlbaum

Catts, H.W. & Kamhi, A. G. (2017). Reading comprehension is not a single ability. *Language, Speech, and Hearing Services in Schools, 48*, 73–76. https://doi.org/10.1044/2017_LSHSS-16-0049.

Clarke, P.J., Henderson, L.M. & Truelove, E. (2010). The poor comprehender profile: Understanding and supporting individuals who have difficulties extracting meaning from text. In J. Holmes (Ed.), *Advances in Child Development and Behaviour*, Vol 39, (pp. 79–129). New York, Boston: Academic Press.

Clarke, P.J., Snowling, M.J., Truelove, E. & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science, 21*, 1106–1116. https://doi.org/10.1177/0956797610375449.

Colenbrander, D., Nickels, L. & Kohnen, S. (2017). Similar but different: differences in comprehension diagnosis on the Neale Analysis of Reading Ability and the York Assessment of Reading for Comprehension. *Journal of Research in Reading, 40*(4)403-419. https://doi.org/10.1111/1467-9817.12075.

Collins, A., Lindström, E. & Compton, D. (2018). Comparing students with and without reading difficulties on reading comprehension assessments: A meta-analysis. *Journal of Learning Disabilities, 51*(2), 108–123. https://doi.org/10.1177/0022219417704636.

Cuetos, F., Rodríguez, B. Ruano, R. & Arribas, D. (2007). *PROLEC-R Batería de evaluación de los procesos lectores, revisada* [Battery of evaluation of the reading processes, revised]. Madrid (Spain): TEA Ediciones.

Cuetos, F., & Suárez-Coalla, P. (2009). From grapheme to word in reading acquisition in Spanish. *Applied Psycholinguistics, 30*(4), 583–601. https://doi.org/10.1017/S0142716409990038

Cutting, L. E. & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10*, 277–299. https://doi.org/10.1207/s1532799xssr1003_5.

Dunn, L. M. & Markwardt, F. C. (1970). *Examiner's manual: Peabody individual achievement test*. Circle Pines, MN: American Guidance Service.

Elosúa, M.R., García-Madruga, J.A., Gómez-Veiga, I., López-Escribano, C., Pérez, E. & Orjales, I. (2012). Habilidades lectoras y rendimiento académico en 3° y 6° de Primaria: aspectos evolutivos y educativos. *Estudios de Psicología, 33*, 207-218. https://doi.org/10.1174/021093912800676411.

Fletcher. J. (2006) Measuring Reading Comprehension. *Scientific Studies of Reading, 10* (3), 323-330. https://doi.org/10.1207/s1532799xssr1003_7.

Francis, D. J., Fletcher, J. M., Catts, H. W. & Tomblin, J. B. (2005). Dimensions affecting the assessment of reading comprehension. In S. A. Stahl & S. G. Paris (Eds.), *Children's reading comprehension and assessment* (pp. 369–394). Mahwah, NJ: Erlbaum.

Francis, D. J., Snow, C. E., August, D., Carlson, C. D., Miller, J. & Iglesias, A. (2006). Measures of reading comprehension: A latent variable analysis of the diagnostic assessment of reading

comprehension. *Scientific Studies of Reading, 10* (3), 301-322. https://doi.org/10.1207/s1532799xssr1003_6.

García, E., Jiménez, J. E., González, D. & Jiménez-Suárez, E. (2013). Reading comprehension difficulties among students in primary education and compulsory secondary education: a study of prevalence in Spanish. *European Journal of Investigation in Health, Psychology and Education, 3*(2), 113-123. https://doi.org/10.30552/ejihpe.v3i2.43

García-Madruga, J.A., Vila, J.O., Gómez-Veiga, I., Duque, G. & Elosúa, M.R. (2014). Executive processes, reading comprehension and academic achievement in third-grade primary students. *Learning and Individual Differences, 35*, 41-48. https://doi.org/10.1016/j.lindif.2014.07.013.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading and reading disability. *Remedial and Special Education, 7*, 6–10.

Gray, S. (2017). Introduction to the clinical Forum: Reading comprehension is not a single ability. *Language, Speech, and Hearing Services in Schools, 48*, 71–72. https://doi.org/10.1044/2017_LSHSS-16-0080

Hagley, F. (1987). *Suffolk Reading Scale*. Windsor: NFER-Nelson.

Hoover, W.A. & Gough, P.B. (1990). The Simple View of Reading. *Reading and Writing, 2*, 127–160. https://doi.org/10.1007/BF00401799.

Hulme, C. & Snowling, M.J. (2011). Children's reading comprehension difficulties: Nature, causes, and treatments. *Current Directions in Psychological Science, 20*(3), 139-142. https://doi.org/10.1177/0963721411408673.

Kamil, M. (2003). *Adolescents and literacy: Reading for the twenty-first century*. Washington, DC: Alliance for Excellent Education.

Kaufman, A.S. & Kaufman, N.L. (1997). *K-BIT. Kaufman Brief Intelligence Test*. Madrid (Spain): TEA.

Keenan, J.M., Betjemann, R.S. & Olson, R.K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12*, 281–300. https://doi.org/10.1080/10888430802132279.

Keenan, J.M., Hua, A.N., Meenan, C.E., Pennington, B.F., Willcut, E. & Olson, R.K. (2014). Issues in identifying poor comprehenders. *L'Année Psychologique, 114*, 753–777. https://doi.org/10.4074/S0003503314004072.

Keenan, J.M. & Meenan, C.E. (2014). Test differences in diagnosing reading comprehension deficits. *Journal of Learning Disabilities, 47*, 125–135. https://doi.org/10.1177/0022219412439326.

LaBerge, D. & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*(2), 293-323. https://doi.org/10.1016/0010-0285(74)90015-2.

León, J., Escudero, I., & Olmos, (2012). *ECOMPLEC: Evaluación de la comprensión lectora* [Reading Comprehension Assessment Test]. Madrid (Spain): TEA Ediciones.

Leslie, L. & Caldwell, J. (2001). *Qualitative Reading Inventory-3*. Addison Wesley Longman: New York.

López-Escribano, C, Elosúa, M.R., Gómez-Veiga, I., García-Madruga, J.A. (2013). A predictive study of reading comprehension in third-grade Spanish students. *Psicothema, 25*, 199-205. https://doi.org/10.7334/psicothema2012.175.

MacGinitie, W. H., MacGinitie, R. K., Maria, K. & Dreyer, L. G. (2000). *Gates–MacGinitie Reading Tests* (4th ed.). Itasca, IL: Riverside.

McLaughlin, M.J., Speirs, K.E. & Shenassa, E.D. (2012). Reading disability and adult attained education and income: Evidence from a 30-year longitudinal study of a population-based sample. *Journal of Learning Disabilities, 47*, 374–386. https://doi.org/10.1177/0022219412458323.

Meneguetti, C., Carretti, B. & De Beni, R. (2006). Components of reading comprehension and scholastic achievement. *Learning and Individual Differences, 16*, 291-301. https://doi.org/10.1016/j.lindif.2006.11.001.

Miller, A.C., Keenan, J.M., Betjemann, R.S., Willcutt, E.G., Pennington, B.F. & Olson, R.K. (2008). Reading comprehension in children with ADHD: cognitive underpinnings of the centrality deficit. *Journal of Abnormal Child Psychology 41*, 473–483. https://doi.org/10.1007/s10802-012-9686-8.

Mullis, I.V.S., Martin, M.O., Foy, P. & Drucker, K.T. (2012). *PIRLS 2011 International Results in Reading*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I. V. S., Martin, M. O., Foy, P. & Hooper, M. (2017). *PIRLS 2016 International Results in Reading*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/pirls2016/international-results/

Nation, K., Snowling, M.J. & Clarke, P. (2005). Production of the English past tense by children with language comprehension impairment. *Journal of Child Language, 32*, 117–137. https://doi.org/10.1017/S0305000904006555.

Nation, K. & Snowling, M. (1997). Assessing reading difficulties: the validity and utility of current measures of reading skill. *British Journal of Educational Psychology, 67*, 359–370. https://doi.org/10.1111/j.2044-8279.1997.tb01250.x

Neale, M. D. (1989). *The Neale Analysis of Reading Ability-Revised*. Windsor: NFER.

Neale, M.D. (1997). *Neale Analysis of Reading Ability: Second Revised British Edition*. London: NFER-Nelson.

Neale, M.D. (1999). *The Neale Analysis of Reading Ability*. (Third edn). Melbourne, Australia: ACER.

Oakhill, J. & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading, 16*, 91–121. https://doi.org/10.1080/10888438.2010.529219.

OECD (2017). *Education at a Glance 2017*: OECD Indicators. OECD Publishing. Available on: http://www.oecd-ilibrary.org/education/education-at-a-glance-2017_eag-2017-en

O'Reilly, T. & McNamara, D. S. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional 'High-Stakes' measures of high school students' science achievement. *American Educational Research Journal, 44*, 161–196. https://doi.org/10.3102/0002831206298171.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4) 357–383. https://doi.org/10.1080/10888430701530730.

Perfetti, C. A. & Hart, L. (2011). The lexical bases of comprehension skill. In D. Gorfien (Ed.), *On the consequences of meaning selection* (pp. 67–86). Washington, DC: American Psychological Association.

Ricketts, J., Sperring, R. & Nation, K. (2014). Educational attainment in poor comprehenders. *Frontiers in Psychology, 5*, 445. https://doi.org/10.3389/fpsyg.2014.00445.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*(2), 143-174. https://doi.org/10.1348/000712603321661859

Wang, J. J. & Kaufman, A. S. (1993). Changes in fluid and crystallized intelligence across the 20- to 90-year age range on the K-BIT. *Journal of Psychoeducational Assessment, 11*(1), 29-37. https://doi.org/10.1177/073428299301100104

Wechsler, D. L. (1992).*Wechsler Individual Achievement Test*. San Antonio, TX: Psychological Corporation.

Wiederholt, L. & Bryant, B. (1992). *Examiner's manual: Gray Oral Reading Test—3*. Austin, TX: Pro-Ed.

Woodcock, R. W., McGrew, K. S. & Mather, N. (2001). *Woodcock–Johnson III tests of achievement*. Itasca, IL: Riverside.

Table 1

Descriptive Statistics (Mean, Standard Deviation, Minimum and Maximum) for All Measures.

Measures	Mean	SD	Min	Max
Reading comprehension tests				
ECOMPLEC (RS)	48.35	6.65	24	61
PROLEC-R (RS)	14.93	1.11	10	16
ACL (RS)	18.85	5.65	5	31
ACL (Decatype)	5.39	2.00	1	10
Decoding measures				
Word reading rate (RS)	62.77	13.80	23	104
Nonword reading rate (RS)	36.00	10.17	16	69
Cognitive measures				
Vocabulary (Verbal intelligence) (RS)	50.29	4.84	38	62
Vocabulary (Verbal intelligence) (SS)	99.92	11.19	75	128
Matrices (Nonverbal intelligence) (RS)	30.11	5.42	17	46
Matrices (Nonverbal intelligence) (SS)	100.88	14.11	66	142
Academic achievement	7.70	1.28	4	9.86

Note. RS = Raw score; SS = Standard score; ECOMPLEC = Reading Comprehension Assessment Test; PROLEC-R = PROLEC-R Reading comprehension subtest for the revised battery of evaluation of reading processes; ACL = Assessment of Reading Comprehension: ACL tests

Table 2

Intercorrelations Among the Three Reading Comprehension Measures, Age, Word reading rate, Nonword reading rate, Vocabulary and Academic achievement.

	1	2	3	4	5	6	7
1. ECOMPLEC							
2. PROLEC-R	.30***						
3. ACL	.67***	.33***					
4. Age	.21*	.14	.17*				
5. Word reading rate	.24**	.16*	.22**	.37***			
6. Nonword reading rate	.27***	.24**	.22*	.45***	.65***		
7. Vocabulary	.45***	.31***	.46***	.01	.11	.07	
8. Academic achievement	.40***	.24**	.42***	.06	.26**	.34***	.32***

Note. ECOMPLEC = Reading Comprehension Assessment Test; PROLEC-R = PROLEC-R Reading comprehension subtest for the revised battery of evaluation of reading processes; ACL = Assessment of Reading Comprehension: ACL tests; *p< .05; **p< .01; ***p< .001

Table 3

Regression Analyses Predicting Reading Comprehension from Age, Decoding Skills and Vocabulary for Different Reading Comprehension Tests

	ECOMPLEC		PROLEC-R		ACL	
Variables	Final β	R^2 Adjusted	Final β	R^2 Adjusted	Final β	R^2 Adjusted
1. Age	.11*	.04*	.06	.01	.10*	.02*
2. Vocabulary	.43***	.23***	.30**	.10**	.44***	.23***
3. Decoding skills	.19*	.26*	.16	.11	.15	.24

Note. N = 139. ECOMPLEC = Reading Comprehension Assessment Test; PROLEC-R = PROLEC-R Reading comprehension subtest for the revised battery of evaluation of reading processes; ACL = Assessment of Reading Comprehension: ACL tests; *p< .05; **p< .01; ***p< .001

Table 4

Prevalence of Reading Profiles and Consistency of Classification for Decoding and Comprehension Skills.

Profile	Percentage of children		
	ECOMPLEC	PROLEC-R	ACL
Poor comprehender			
decoding ≥ (mean - 1.5 SD); reading comprehension > (mean - 1.5 SD)			
Word reading rate	6	10	6
Nonword reading rate	7	9	6
Poor reader			
decoding <(mean - 1.5 SD); reading comprehension <(mean - 1.5 SD)			
Word reading rate	2	1	1
Nonword reading rate	2	2	1
Poor decoder			
decoding <(mean - 1.5 SD); reading comprehension ≥ (mean - 1.5 SD)			
Word reading rate	4	4	5
Nonword reading rate	5	4	6
Successful reader			
decoding ≥ (mean - 1.5 SD); reading comprehension ≥ (mean - 1.5 SD)			
Word reading rate	88	83	88

Nonword reading rate	86	84	87

Note. ECOMPLEC = Reading Comprehension Assessment Test; PROLEC-R = PROLEC-R Reading comprehension subtest for the revised battery of evaluation of reading processes; ACL = Assessment of Reading Comprehension: ACL tests.

The percentage of children in each reading were calculated using whole numbers following Colenbrander et al's study (2017).

Table 5

Consistency of Classification for All the Sample and for Each Reading Profile Across Comprehension Tests Using Word and Nonword Measures

Profile	Consistency (percentage)
All the sample	
Word reading rate	79
Nonword reading rate	81
Poor comprehender profile	
Word reading rate	8
Nonword reading rate	4
Poor reader profile	
Word reading rate	0
Nonword reading rate	20
Poor decoder profile	
Word reading rate	43
Nonword reading rate	62
Successful reader profile	
Word reading rate	82
Nonword reading rate	81