ASSESSMENTS OF READING COMPREHENSION: DIFFERENT MEASURES, DIFFERENT EXPLANATORY VARIABLES AND ACHIEVEMENT

LUISA ARAUJO*, JOSÉ MORAIS**, & ARMANDA COSTA

**Joint research centre of the European commission, ispra, Italy, and Instituto superior de educação e ciências, Lisbon, Portugal,*
*** University Libre de Bruxelles,*
**** University of Lisbon*

Abstract
Assessments of Reading Comprehension: Different Measures, Different Explanatory Variables and Achievement Abstract This study investigates the reading achievement of 69 Portuguese fourth graders on national (NA) and international (PIRLS) assessments and its relation to reading skills in the following areas: phonological awareness, word recognition, fluency, accuracy, and types of oral reading errors. Data analysis was conducted using regression and t-tests for dependent samples. Results indicate that in the NA reading accuracy and vocabulary knowledge predict reading comprehension. In PIRLS, vocabulary knowledge and non-meaning changing errors explain the variance in reading scores. Performance in the NA was significantly higher than in PIRLS and, in this test, students scored significantly lower on questions that require interpreting and evaluating text than on literal questions. Differences in test design and implications for monitoring educational outcomes are discussed.

Keywords: educational evaluation, fourth-grade, international surveys, reading comprehension

Araújo, L., Morais, J., & Costa, A. (2013). Assessments of Reading Comprehension: Different Measures, Different Explanatory variables and Achievement. L1-Educational Studies in Language and Literature, vol.13, *p. 1-21. http://dx.doi.org/10.17239/L1ESLL-2013.01.04*
Corresponding author: Luísa Araújo, Joint Research Centre/Instituto Superior de Educação e Ciências email: luisabaraujo@gmail.com

1. INTRODUCTION

Reading comprehension can be defined as …"an active and complex process that involves understanding written text; developing and interpreting meaning; and using meaning as appropriate to type of text; purpose and situation" (National Assessment Governing Board, 2007, p. 2). At present, international evaluations of reading achievement assume that by grade four students have acquired the basic skills needed to comprehend what they read. In particular, the Program for International Reading Literacy Study (PIRLS), conducted by the International Association for the Evaluation of Educational Achievement (IEA), designed to assess reading at the fourth grade level on a five-year cycle, is conceptually close to the reading stages proposed by Jeanne Chall (1996). It seeks to capture students` progress in moving from the first stage of learning to read to the subsequent reading to learn stage.

The PIRLS framework (Mullis, Kennedy, Martin, & Sainsbury, 2006) summarizes the goals of both PIRLS and PISA as follows: "While PISA is concerned with the literacy needs of students as they make the transition from the world of school to the world of work, PIRLS addresses progress at the equally important stage when students move from learning to read to reading to learn" (p. 102). The goal of these international assessments is to measure students´ ability to comprehend what they read, both to fulfill personal goals and to participate in society, and their goal is to provide participating countries trend information about the reading literacy performance of their students (Shiel & Eivers, 2009). In PIRLS 2006, 45 countries participated and in 2011 55 countries participated.

The fact that the fourth grade constitutes the first moment for international performance monitoring is supported by research that shows that failing to learn to read fluently with good comprehension before the third or fourth year of schooling may result in life-long problems in learning new skills. In the words of Keith Stanovich, "This is because children who fall behind in reading, read less, increasing the gap between them and their peers. Later, when students need to "read to learn" (where before they were learning to read) their reading difficulty creates difficulty in most other subjects. In this way they fall further and further behind in school, dropping out at a much higher rate than their peers" (Adams, 1990, p. 59).

2. PREDICTING READING ABILITY

Decoding ability has long been recognized as the basis for successful reading development in alphabetic languages, regardless of the specificities of different orthographic systems. Decoding refers to the mechanism of assigning a phoneme to each grapheme. This is followed by phonological recoding, the fusion or integration of the recovered phonemes in order to form each of the successive syllables (see Adams, 1991 for a review). In deep or opaque orthographies like English there are

many multi-letter graphemes (letter combinations that represent one phoneme or speech sound as in boat). In transparent writing systems such as the Finnish there is an almost perfect one to one correspondence between graphemes and phonemes and a simple syllable structure or a predominance of open CV syllables with few initial or final clusters (Seymour, Aro, & Erskine, 2003). In the early stages of learning to read these written code differences seem to result in differing rates of reading development. Specifically, childrenlearning to read in English seem to take longer to achieve the same level of decoding ability of the majority of their European counterparts. French, Portuguese, and Danish first graders seem to take a little longer than German or Spanish ones to acquire basic decoding skills because they have to learn slightly more irregular languages (Seymour et al., 2003). Research studies with other orthographic codes also suggests that "orthographies that represent pronunciation encourage faster learning for reading aloud and the more transparently they do this, the faster the learning rate and the more they encourage lexical access via phonology" (Ellis, Natsume, Stavropoulou, Hoxhallari, Van Daal, Polyzoe, Tsipa, & Petalas, 2004, p. 458).

Indeed, comparative international research has recently shown that, although there are weaker correlations between phonological awareness and reading ability in transparent orthographies, the predicting power of phonological awareness is universal (Ziegler, Bertrand, Tóth, Csépe, Reis, Faísca, Saine, Lyytinen, Vaessen, & Blomert, 2010). Thus, phonological awareness is more predictive of reading ability in opaque languages, but it remains the main factor associated with reading performance in all the alphabetic systems studied to date.

Good decoding skills, based on phonological processing, enable readers to form stable orthographic representations which make word identification automatic and effortless. In turn, this automaticity makes it easier for readers to comprehend what they read. Indeed, several decades of research have confirmed LaBerge and Samuels´ (1974) model of reading showing that "slow reading of words consumes the limited consciousness available for processing text, with the result that no consciousness is left over for understanding what is read" (Pressley, Gaskins, & Fingeret, 2006, p. 54). In a similar vein, Perfetti´s (1992) verbal-efficiency account of reading postulates that fast word-identification skills serve as the foundation for text comprehension. Evidence from research with students with reading difficulties, in particular, confirms that the great majority of those students have not developed adequate decoding and word recognition abilities (Stanovich, 2000). Recent psycholinguistic research confirms that there are two routes leading to the correct pronunciation of words (Coltheart, Curtis, Atkins & Haller, 1993; Perry, Ziegler & Zorzi, 2010). On the one hand, the sub-lexical or grapheme-phoneme conversion route allows us to identify the phonological form of all sorts of words, including words we have never seen before (e.g. fass). The direct access route, on the other hand, allows us to recognize known words by using orthographic information stored from previous encounters with the words. It accounts for the lexical processing of irregular words, those where phoneme-graphemes conversion rules do

not apply, and of homographs (e.g. reading tear by analogy with dear) (Morais, 2003).

2.1 Assessing Reading Comprehension

As reading ability develops, other factors such as vocabulary knowledge become associated with reading comprehension. For example, Sénéchal, Ouellette, and Rodney (2006) found that young children's vocabulary knowledge, acquired from parental book reading, predicts reading in grade three. In contrast, Chall and Jacobs´ (2003) research shows that low achieving fourth graders begin to exhibit a vocabulary lag, namely difficulty in defining less common words, and that their reading comprehension scores begin to decline significantly in grade six. Conclusions from the National Reading Panel meta-analysis further corroborate this indicating that recreational reading is a privileged way to learn reading skills implicitly. Students with strong reading habits consolidate orthographic representations of words theyrepeatedly encounter in print and constantly learn new vocabulary (Nagy & Scott, 2000; Kamil, 2004).

As a result of repeated recreational and teacher-guided reading experiences, students also gain world knowledge and develop inference skills (Guthrie, Wigfield, & VonSecker, 2000). Research suggests that poor comprehenders may have difficulty at the level of inferences and in integration of information (Cain, Oakhill, Barnes, & Bryant, 2001; Yuill & Oakhill, 1991). In fact, international studies suggest that low achieving students evidence more weaknesses in integrating world knowledge with text-based information and in critically evaluating text content than in retrieving explicitly stated text information. In the PIRLS 2001 study "… it was discovered that, for the passages in the assessment, students at the lower quarter benchmark demonstrated the most success on items requiring retrieval of explicitly stated details from the various literary and informational texts" (Mullins, Martin, & Gonzalez, 2004, p. 24). However, in the subsequent PIRLS 2006 assessment the same pattern of achievement was not observed.

In sum, there is empirical evidence supporting the idea that reading is constrained by bottom-up processing, but that other factors also explain reading comprehension. Some researchers have measured word-level processing ability in terms of reading speed and decoding accuracy and found that it explains more of the variance in reading attainment in the early grades than in subsequent years (Catts & Hogan, 2003; Francis, Fletcher, Catts, & Tomblin, 2005; Spear-Swerling, 2005). Others have found a significant contribution of phonological awareness until grade four while others have found such influence only until grade two (Patel, Snowling, & Jong, 2004¸ Ziegler et al., 2010). As Stahl (2004) states, with respect to comprehension measures, "oral reading accuracy may be important only in the early grades with other factors such as vocabulary knowledge and comprehension strategies use becoming important later on" (p. 189).

The 2002 NAEP oral reading study confirmed that both oral reading accuracy and oral reading rate are related to reading comprehension (U.S Department of Education, 2005). These findings, however, contradict previous results by Pinnell, Pikulski, Wixson, Campbell, Gough, and Beatty (1995) who found that accuracy was not significantly related to comprehension, but that instead the number of significant miscues was. Similarly, research clearly shows that word-level fluency is a precursor to fluent reading of connected text, as measured by the number of words read per minute, and that the latter enables reading comprehension, but the relation between fluency and comprehension is not fully understood (Pressley, Gaskins, & Fingeret, 2006).

Stanovich´s (1980) interactive-compensatory model, for example, predicts that context can compensate for inefficient word processing skills. More specifically, it "compensates for poor readers´ slow print processing when it delivers top-down information about a word's identity before bottom-up processing has concluded" (Jenkins, Fuchs, Van Den Broek, Espin, & Deno, 2003, p. 720). Nevertheless, as Adams points out "…research has taught us that written text is accessible, and thus permits learning, only if the reader or listener already knows the vast majority of words from which it is constructed. Indeed, research indicates that reading with comprehension depends on understanding at least 95% of the words of the text" (Adams, 2009, p.172). This understanding involves vocabulary knowledge as well as knowledge of the world in general. Research suggests that some poor comprehenders may have reached appropriate fluency and accuracy levels for their respective age/grade and have adequate word and world knowledge, but exhibit a specific difficulty in answering questions that require inference skills (Cain &Oakhill, 1999; Perfetti, Landi, & Oakhill, 2005). For example, in understanding text-level inferences that require attending to referential links as in John took the drink out of his bag. The milk was very refreshing, we can logically infer that the drink was milk.

Other types of textual information require pragmatic inferences, which involve the consideration of a situation model based on knowledge of the world. This happens, for example, when a reader needs to gather information from different parts of a text to infer the setting of a story or when knowledge of a particular word makes an inference possible. For instance, when we read a statement such as "dressed in silk", depending on the overall tone of the passage, we can infer that the person is well off (Giasson, 1990).

2.1.1 Large-scale Assessments of Reading Literacy Achievement

PIRLS defines reading literacy as "the ability to understand and use those written language forms required by society and/or valued by the individual. Young readers can construct meaning from a variety of texts. They read to learn, to participate in communities of readers in school and everyday life, and for enjoyment" (Mullis et al., 2006, p. 103).

In addition to considering two reading purposes – for literary experience and for information – the PIRLS framework also considers reading processes. Included in the framework are "... four specific processes of reading comprehension, which vary in the degree of inference or interpretation required and in the focus on text content or structural features of the text" (Sainsbury & Campbell, 2003, p. 16). They are 1) *Focus on and Retrieve Explicitly Stated Information*, which requires students to locate and understand relevant information or ideas that are explicitly stated in text, 2) *Make Straightforward Inferences*, or move beyond surface meaning to make straightforward, text-based inferences, 3) *Interpret and Integrate Ideas and Information*, whereby students have to draw on their understanding of the world, experience, or other knowledge to find connections between ideas and information in the text, and 4) *Examine and Evaluate Content, Language, and Textual Elements*, which requires critical consideration of the text in terms of reflecting on and evaluating text structure and content, language use, literary devices, and the author's perspective and craft (Mullins et al., 2006). Clearly, PIRLS adopts a framework for assessing reading comprehension processes that incorporates a view of reading literacy as an interactive process requiring different skills in processing text information.

The National Assessment of Educational Progress (NAEP) conducted in the United States every two years to measure fourth graders reading performance contemplates similar purposes and processes. The most recent PIRLS 2006 and NAEP 2007 report on analogous fourth grade populations and comparisons suggest that, although similar, the NAEP test demands more complex reading skills (U.S Department of Education, 2003). This is because passages are longer in NAEP than in PIRLS, averaging 1000 words in the former and 547 words in the latter. In addition, when compared with PIRLS, it seems that "the NAEP texts are more complex, with more embedded clauses and a more complex syntactical structure" (U.S Department of Education, 2003, p. 11). Indeed, according to conservative readability statistics which take into account the average number of sentences and syllables per 100 words, NAEP passages are written at a 7^{th} grade level, whereas PIRLS passages are written at a 5^{th} to 6^{th} grade level (U.S Department of Education, 2003). Finally, NAEP tests include more questions that require interpretation than PIRLS and do not ask students to locate information that matches verbatim the questions asked. These findings strongly suggest that NAEP is more demanding than PIRLS (U.S Department of Education, 2003).

Rationale for the Study

According to Foorman (2009), tests are increasingly seen and used to evaluate school systems and to measure progress. However, she asks: "How does one know if a fourth-grade passage on the National Assessment of Educational Progress (NAEP) – referred to as the "the nation's report card" – is equal in difficulty to a fourth-grade passage on a state test, or if the fourth-grade passages on two states´

assessments are comparable?" (Foorman, 2009, p. 231). Similarly, how does one know if national and international assessments are comparable? And could bottom-up processes and other reading skills such as fluency, accuracy rate and vocabulary knowledge have a differentiated impact on reading achievement depending on the type of reading test fourth graders take?

Given that existing research suggests that different explanatory factors may be related to different measures of reading comprehension and that international assessments seek to inform national educational policies, research that investigates possible differences in reading comprehension measures and their associated explanatory variables seems warranted. In particular, when one of the goals of international assessments is to inform national educational policies, it seems appropriate to examine exactly what they measure and how they compare to national tests.

Purpose

The purpose of this study was to investigate how PIRLS and Portuguese national assessments compare with respect to measuring reading comprehension at the fourth grade level. In the 1991 Study of Reading Literacy, a precursor of PIRLS, Portuguese fourth graders performed below the international average in reading comprehension (Elley, 1992). The results of this study showed that there were no statistically significant differences in Portuguese students' performance according to the three reading domains tested: narratives, expository texts and documents. We chose to compare fourth graders reading achievement in PIRLS and in the Portuguese national assessment (henceforth NA) because such comparison may shed light on how different reading demands may call upon different reading processes and require different abilities. More specifically, we aimed at answering the following questions:

1) Are the contributions of phonological awareness, word recognition/decoding, vocabulary, fluency and accuracy to reading connected text, different for different reading measures?
2) How do the two reading tests compare in terms of accounting for reading comprehension?

3. METHOD

3.1 Materials

For both the NA and the PIRLS measures of reading comprehension reading assessment passages used in 2001 were selected. For the latter, a test was selected from one of the booklets used in 2001 to assess literary experience. Both tests included narratives from well known children's authors and asked approximately the same number of questions -14 for PIRLS and 13 for the NA - mostly multiple choice and constructed responses. The latter require that student write their own response to the questions, rather than checking the right answer from four possibili-

ties. However the NA also included sequencing, true/false and sentence completion questions. Both required about the same testing time, 35 minutes for PIRLS and 40 minutes for NA. To control for this small difference, students were given 40 minutes to take each test. Since Portugal did not participate in PIRLS 2001 or 2006, two scholars with a degree in translations studies translated the English version of this international assessment, discrepancies related to nuances in meaning were noted and a final version agreed upon. Table 1 shows, for the two instruments and for the original PIRLS passage in English, the text metrics that are usually considered in readability formulas. Both the translated version of PIRLS and the NA have very similar metrics, with the translated PIRLS version registering 20 words more than NA.

Table 1. Text metrics

	Nr. words	Nr. Sentences	Nr Paragraphs	Mean Ratio Sentences/ Paragraph	Mean . Ratio Words/ Sentence	Mean Ratio Syllables/ Word
PIRLS 2001 original	524	44	21	2,0	12	1,2
PIRLS 2001 Portuguese Translation	476	41	25	1,7	11,6	1,9
NA test 2001	453	42	23	1,8	10,7	2,0

For the original PIRLS passage, the readability statistics available in Microsoft Word 2003 point to a *Flesh Reading Ease* score of 86 and a 3.6 Flesch Kincaid grade level. Although slightly higher, these figures are consistent with the findings of a comparative analysis between the American National Assessment of Educational Progress (NAEP), also at the 4th grade level, and PIRLS that considered all booklets used in the PIRLS 2001 study; four narrative and four informational (NCES, 2003). In that comparative analysis, the *Flesh Reading Ease* score for all the PIRLS passages was 81.7 and the Lexile analysis, which takes into account semantic difficulty and syntactic complexity, indicated a 3rd to 4th grade readability level (NCES, 2003). Thus, the PIRLS passage used in this study was very close to the average for all the PIRLS 2001 passages with respect to reading level (NCES, 2003).

For the Portuguese translation of the PIRLS passage, the *Flesh Reading Ease* score was 35. However, given that the Reading Ease score equation includes dividing syllables by words, a text's difficulty in a language like Portuguese with a large

number of multisyllabic paroxitone words increases considerably when compared to an English text (Cardoso-Martins, 1995). In order to account for these differences, the Huerta formula, developed for Spanish, a language similar to Portuguese, was used to find the readability level of the translated PIRLS passage and of the NA text (Fernández, 1959). Accordingly, the PIRLS translation resulted in a readability level of 81.

Similarly, using the Huerta formula the readability level of the NA passage rose from 41, in the *Flesh Reading Ease*, to 80; a score within the same "Easy" interval (80-90) as that found for the original PIRLS text by Rohl Dahl and for the translated version of this passage. According to a Portuguese study, the selected NA passage also corresponds to the average reading difficulty for narrative texts used in yearly national assessments from 2000 to 2008 (Correia, 2010).

With respect to comprehension processes, the differences between the two tests are more pronounced. Table 2 shows the percentage of questions by type of comprehension process and compares the two tests used in this study with all the narrative passages used in PIRLS 2001 (4 booklets).

Table 2. Percentage of questions according to the PIRLS Reading Comprehension Framework

Reading Comprehension Processes	All PIRLS 2001 Booklets	PIRLS 2001: The up-side down mice by Roald Dahl	NA 2001: *The Oriana Fairy* by Sophia de Mello Bryner
Focus on and Retrieve Explicitly Stated Information	20%	21%	40%
Make Straightforward Inferences	40%	29%	40%
Interpret and Integrate Ideas and Information	25%	29%	20%
Examine and Evaluate Content, Language, and Textual Elements	15%	21%	0%

The relative weight of the types of comprehension processes in the Rohl Dahl passage used in 2001 is more akin to the distribution adopted for all booklets in 2006 and 2011 than in 2001 (Mullins et al., 2006). Specifically, this passage reflects a 50/50 ratio between the first two text-based processes of reading; retrieve information and make straightforward inferences and the other two more interpretative and knowledge-based processes; integrate ideas and evaluate content, language and textual elements. Clearly, the Portuguese test places an overwhelming emphasis on the first two reading comprehension processes.

Both assessments included multiple choice questions, worth one point each, and constructed responses where students had to write their responses and which were worth anywhere from one to five points. In addition, the NA included 1 sequencing question, 1 true/false question and 1 sentence completion question. There were a total of 13 questions in this test compared to 14 in PIRLS. Table 3 shows the distribution of the types of questions according to the different comprehension processes.

Table 3. Types of questions according to the PIRLS Reading Comprehension Framework

Reading Comprehension Processes	PIRLS 2001: The upside down mice by Roald Dahl	NA 2001: *The Oriana Fairy* by Sophia de Mello Bryner
Focus on and Retrieve Explicitly Stated Information	2 multiple choice (#2, 5) 1 written construction (#10)	Multiple choice (#1,3, 8) 1 written construction (#2, 7, 11) 1 fill in the blank (#9) Sequencing (#10)
Make Straightforward Inferences	3 multiple choice (#1, 3, 9) 1 written construction (#7)	1 True/False (#6) 1 written construction (#13)
Interpret and Integrate Ideas and Information	4 written construction (#4, 6, 11, 12)	2 Multiple choice (# 4, 12), 1 fill in the blank (# 5)
Examine and Evaluate Content, Language, and Textual Elements	2 multiple choice (8, 13) 1 written construction (14)	-

3.2 Participants

Participants were 69 fourth graders in three intact, self-contained classrooms in a school located in a suburban middle-class town in Portugal. They were tested at the beginning and at the end of the school year. Students receiving special education services were not included in the sample, which had a nearly equal number of males and females. The three teachers assigned to these classes, one male and two females, taught literacy using the same basal series and consisted of the existing fourth population in the school that agreed to participate in the study.

3.3 Measures

We used two dependent measures and several independent measures. The first dependent measure, the Portuguese NA test was used twice; at the outset of the study, in October, to evaluate whether there were differences in reading ability among the classes and in June, at the end of the study, at which time the PIRLS test was also used as a dependent measure of reading comprehension. A one-way analysis of variance done for the national test revealed no significant differences between the three classrooms, $F(2, 66) = .083$, $p = .921$, in October at the beginning of the school year. The independent measures selected included variables that have been shown to explain reading ability.

3.4 Independent Measures

3.4.1 Pseudo-word reading

In order to assess the predictive value of grapho-phonological decoding, we used a pseudo-word test adapted from a Brazilian one (Scliar-Cabral, 2003). Two words were eliminated from the original 38 due to orthographic differences between Brazilian Portuguese and European Portuguese (e.g., words containing the accent mark ¨ in letters like ü). The remaining words reflected a wide range of orthographic patterns in Portuguese; from different diphthongs to different vowel and consonant digraphs. Students were asked to read the list of 36 two-syllable pseudo-words presented in "8.5 x 11" sheets with a 14 inch font, double-spaced. This procedure was timed by having the researcher slide a card with an open square where only one word at a time could be seen following a 10 seconds interval for each word. One point was given for each pseudo word read correctly.

3.4.2 Word reading

Because there is a word meaning aspect in lexical decisions, word reading may have a better predictive value over pseudo-words in reading comprehension (Gijsel, van Bon & Bosman, 2004). Thus, a word recognition test consisting of a list of 20 words constructed according to varying degrees of frequency and different orthographic patterns was used. The European lexical database, Portulex, which includes words that appear in basal readers from first through fourth grade, was used (Gomes & Castro, 2003). From this database, words that reflected different frequency levels (presence ranging from 5 to 75 percent of the corpus), different levels of orthographic complexity (including regular and irregular words), and a range of syllabic structures (CVC, CV, V, VC, CCV, CCVC) were selected (www.portulex.net, 2010). Students were asked to read the list of 20 words presented in "8.5 x 11" sheets with a 14 inch font, double-spaced. This procedure was

timed and the test scored in the same way as the pseudo-word reading test (see Appendix I for word list).

3.4.3 Vocabulary knowledge

A standardized version of the Spanish Wisc-R vocabulary sub-test was translated to Portuguese and used to access students´ vocabulary knowledge, a variable recognized for its impact on reading comprehension. The instructions given to the students followed standard manual procedures calling for the definitions of the words and the students´ responses were written down by the researcher and audio recorded. The coding of the answers also followed standard manual instructions.

3.4.4 Oral reading fluency

This measure of text fluency was obtained by using a 412 words narrative text within the same reading level (Huerta = 85) as the ones used for testing reading comprehension. We shall refer to this test as the *Read-aloud Test*. To assess fluency as a time-based measure (reading rate) we calculated the total number of seconds students took to read the entire passage. First, we subtracted the number of words misread, inserted or skipped from the total number of words in the passage to get the number of words read correctly. Self-corrections were not counted as errors. Then we divided the number of words read correctly by the number of seconds it took to read the passage and multiplied the total by 60 to get a measure of words read correctly per minute (WCPM).

3.4.5 Oral reading accuracy

Reading accuracy was determined by considering the total number of errors students made while reading the *Read-aloud Test* and the 412 words in the text (total words read – total errors) / total words read x 100 = Accuracy rate). Thus, reading accuracy includes all errors children made, except self-corrected ones. Self-corrections can be expected to impact reading rate, which is captured in the reading fluency measure.

3.4.6 Number of meaning-changing errors

Self-corrections during the *Read-aloud Test* were not counted as errors and the remaining miscues were classified according to three types of errors: 1) insertions, 2) omissions, and 3) substitutions. From the total of these combined categories, the number of meaning changing errors alone was entered as another independent variable. For example, if a student substituted a word by its equivalent diminutive or read an indefinite article instead of a definite one and these miscues did not significantly alter the meaning of the passage they were not counted as errors.

3.4.7 Number of errors that do not affect meaning

Following the same rationale adopted to score the number of meaning-changing errors, self-corrections during the *Read- aloud test* were not counted as errors and the remaining errors were classified according to three types of errors: 1) insertions, 2) omissions, and 3) substitutions. From the total of these categories, the number of errors that did not alter meaning was entered as another independent variable.

3.5 Procedure

The reading comprehension tests and the independent measures were administered by the first author to students in the three classrooms. While these tests were collectively administered to students the other measures were administered individually so that the author could record the individual responses to specific stimuli such reading words from a list. In October, students in the three classrooms took the NA test and between October and December they were tested on the Pseudo-word reading, Word Reading test, Wisc-Vocabulary, and took the Read-aloud test, which captured fluency and accuracy measures as well as the number of meaning and non-meaning changing errors. In June, at the end of the school year, students in the three classrooms took the NA again and the PIRLS test.

3.6 Analysis Approach

After the one-way analysis of variance showed no differences between classes at the outset of the school year ($F(2, 66) = .083$, $p = .921$) a forward linear regression analysis was applied to determine which independent variables were predictors of reading comprehension for both the PIRLS test and the NA test. To evaluate how the two dependent measures (NA and PIRLS) compared to one another, a Pearson correlation was calculated and a t-test was used to determine if there were significant differences in the achievement scores between the two tests. In addition, a t-test was also used to evaluate if there were differences in students´ scores between the NA taken at the beginning and at the end of the school year. Finally, additional one tailed t-tests for dependent measures were calculated to determine whether students scored higher on more basic processes of comprehension than in the ones requiring a higher level of interpretation in both the PIRLS and the NA test. For this purpose, we followed the same categorical criteria as that reported in the PIRLS international report. That is, we grouped the explicit and the straightforward inferencing questions in one category and compared the score students obtained in this category versus the score they obtained in another category encompassing the interpretative and evaluative questions. SPPS version 18.0 was used for all statistical analysis.

4. RESULTS

Table 4 shows the descriptive statistics for all tests used. The scores reveal that none of the tests produced ceiling or floor effects. Skewed distributions were found for the dependent measures and for vocabulary, oral fluency and non-meaning changing errors. The Mann-Whitney-Wilcoxon test was used to normalize the distributions, but the transformed values had no effect on subsequent analyses. Therefore, we report the analyses on untransformed scores. The range of scores for the independent and dependent variables is quite wide and, on the former, the smallest variation and the highest students´ scores occur in the Word Reading test and on the Oral Reading Accuracy measure. The widest variation is found in the Oral reading measure and on the Vocabulary test. With respect to the dependent variables, students´ average performance is higher on the NA than on PIRLS.

Table 4. Means, standard deviations, and ranges for independent and depend variables

Test	Mean	*SD*	Range
Pseudo-word reading (max. 36)	30.41	3.2	21-36
Word reading (max. 20)	18	2.37	7-20
WISC-III Vocabulary (raw score)	24.93	5.05	15-38
Oral reading fluency (words correct per minute)	90.26	22.16	43-159
Oral reading accuracy (% of words read correctly)	97.96	1.32	94-100
Meaning-changing errors	3.07	3.66	0-21
Errors that do not affect meaning	5.48	2.80	1-13
NA comprehension test/Beginning of the year (max. 19)	14.86	3.09	6-19
NA comprehension test/End of the year (max. 19)	14.96	2.91	5-19
PIRLS comprehension test (max. 17)	11.06	2.82	3-17

Table 5 lists the correlations among the predicting measures. The values reflect a range of low to strong relationships. Similarly, Table 6 lists the correlations between the dependent measures and the predicting variables and shows that the majority of the relationships reflect moderate correlations.

Table 5. Correlations among Predictors

Variable	1	2	3	4	5	6	7
Pseudo-word reading	__						
Word Reading	.55**	__					
Vocabulary knowledge	.44**	.39**	__				
Oral Reading Fluency	.53**	.57**	.40**	__			
Oral Reading Accuracy	.64**	.62**	.31*	.66**	__		
Meaning-changing errors	-.57**	-.62**	-.25*	.-57**	-.84**	__	
Errors that do not affect meaning	-.45**	-.36**	-.21	-.54**	-.77**	.35**	__

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

Table 6. Correlations among predictors and comprehension tests

Independent Variables	NA test	PIRLS Test
Pseudo-word reading	.51**	.49**
Word Reading	.50**	.50**
Vocabulary knowledge	.41**	.59**
Oral Reading Fluency	.57**	.38**
Oral Reading Accuracy	.64**	.43**
Meaning-changing errors	-.62**	-.44**
Errors that do not affect meaning	-.43**	-.30*

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

The regression analysis shows that a few independent variables explain a considerable amount of the variance in students´ achievement and that there was an acceptable degree of independence between the predicting variables (the Durbin-Watson test indicated a value of 2.21 for the NA regression model and of 2.29 for the PIRLS model). For the NA, the best fitting model indicates that 44% of the variance in students´ achievement is explained by oral reading accuracy and by vocabulary knowledge, with accuracy accounting for 7% and vocabulary accounting for 5% of the variance in the comprehension scores. For PIRLS, the regression model explains 42% of the variance. Vocabulary knowledge appears with the strongest relationship with achievement followed by the number of significant miscues, the first accounting for 7% of the variance and the second for 6%.

Thus, whereas vocabulary knowledge explains achievement in the two reading tests the number of meaning- changing errors affects students´ achievement only in PIRLS and oral reading accuracy explains achievement only in the NA test. Table 7 shows the different relationships between the two dependent measures and their respective predicting variables. The positive association between vocabulary knowledge and comprehension, although present in both assessments, is stronger in PIRLS, as the coefficients shows.

Table 7. Prediction of Reading Comprehension: Regression Analysis for two tests

	NA		PIRLS	
Regression Model:	R^2	β	R^2	β
Predicting reading comprehension	.439		.418	
Oral Reading Accuracy	.073	.564**		
Number of Meaning-changing errors			. 060	-.310*
Vocabulary	.047	.237*	. 071	.508**

** significant at $p < .000'$; * significant at $p < .05$

The correlation coefficient (r = .67) shows that the two reading assessments are moderately correlated. The t-test indicates that students performed significantly better in the NA than in the PIRLS test (t = 13.877, p<.001). Their mean score was 14.96 out of 19 points in the NA, compared with 11.06 out of 17 in PIRLS (in percentage 78.7% and 65.1%, respectively). The largest difference between students' scores in the two tests occurred in the middle and at the high-end of the distribution, with more students scoring at the 50th and 75th percentiles in the NA than in PIRLS. The median in the NA was slightly higher (.351) than the median in the PIRLS assessment (.340).

The results also show that students´ performance on the NA assessment did not improve significantly from the beginning to the end of the school year (t = 0.337, p. = 0.369). Finally, on the PIRLS test, students scored significantly higher on retrieval and straightforward inference questions than on interpreting and evaluating questions (t = 16.469, p < 0.001). On the contrary, in the NA assessment students scored significantly higher on the higher level comprehension processes than on the retrieval and straightforward inferencing questions (t = 4.88, p < 0.001).

5. DISCUSSION

This study indicates that the two reading comprehension tests correspond to different reading difficulty levels related to different loads in the component abilities that explain reading ability. As such, it extends previous research indicating that the contribution of different components may depend on the type of test used to measure reading comprehension. Also, unlike previous findings suggesting that the NAEP assessment used in the US, when compared to the PIRLS assessment, is more demanding, this study shows that the NA, the Portuguese national test, is considerably easier than PIRLS. Indeed, students scored higher on NA than on PIRLS. Nonetheless, it remains difficult to determine which characteristics of the passages are associated to the difference in comprehension attainment in the two assessments. Much care was given to selecting texts that were good prototypes of each reading assessment and that had identical readability levels. However, as was the case in Andreassen & Braten´s (2009) study, the choice of authentic assessments students are likely to encounter in actual test situations makes it difficult to attribute different predicting measures to a particular dimension of the tests.

For example, The Portuguese translation of PIRLS has slightly more words than the NA passage as well as a higher average of number of words per sentence, which can reflect a higher degree of syntactic complexity. PIRLS included proportionally more higher - level comprehension questions than the NA test and this appears to be the most relevant differential characteristic since students´ scores in PIRLS were significantly lower on these questions than on the retrieve and straightforward questions. However, the opposite occurred in the Portuguese test; students scored higher on interpretative questions than on the ones that only required retrieval of textual information. This may be related to the fact that only 20% of the questions belonged to the *Interpret and Integrate Ideas and Information category* and to the absence of questions in the *Examine and Evaluate Content, Language, and Textual Elements* category. In addition, whereas the PIRLS assessment included written construction types of questions to test students' ability to interpret the text the NA assessment tested this ability by including multiple choice and fill in the blank types of questions. As previous studies have indicated, "…commonly used tests of reading comprehension do not necessarily tap the same array of cognitive processes and may be influenced to differing degrees by particular skills that can influence comprehension" (Cutting & Scarborough, 2006, p. 294).

None withstanding, this study corroborates previous findings indicating that vocabulary is a strong predictor of reading ability (Chall & Jacobs, 2003), lending support to the view that it is often a "misunderstood giant" (Sénéchal et al., 2006). With respect to other skills, it is reasonable to expect differences in text comprehension to be related, among other things, to different dimensions of word reading when the comprehension demanding level is also different. The fact that the number of meaning changing errors explained achievement in PIRLS and instead Oral Reading Accuracy explained variance in the NA suggests that meaning changing

errors may be more predictive of reading ability in more complex reading tasks. As Catts (2009) argues, we should not underestimate the complexity of reading comprehension. Clearly, "Comprehension is not a skill like word recognition that can be mastered in a relative short time, but rather a collection of knowledge and processes that takes many years to acquire" (Catts, 2009, p.178). Similarly, readability statistics are useful indicators of reading difficulty levels but do not take into account the vocabulary knowledge and world knowledge students need to interpret text.

The other potential explanatory factors included in the regression analysis, namely isolated word reading and pseudo-word-reading accuracy (the latter provides an indirect reflection of phonological awareness abilities), did not predict students performance in either one of the reading assessments. Thus, unlike findings by Patel et al. (2004) for English and Dutch and by Vaessen, Gerretsen and Blomert (in press) for languages with a similar degree of transparency, we did not find a substantive contribution of grapho-phonological decoding or, indirectly, of phonological awareness in grade four. The nature of the dependent measures we used may account for this different finding. Specifically, to explain reading ability beyond the first grade, word reading fluency measures (correct words/time) of isolated words have been used as dependent variables (Ziegler et al., 2010). Alternatively, cloze reading comprehension tests where students have to supply missing words in very short passages have been used when phonological awareness has been found to be associated with this ability (Caravolas, Volín & Hulme, 2005). This was not the case in the present study which was designed to investigate the relationship between predictors of reading ability and comprehension of considerably longer, connected text. However, if a timed measure of pseudo-words read in isolation had been used, it is possible that individual differences in word reading fluency might also account for reading comprehension. Thus, not accounting for a time measure may have limited our results.

Another potential limitation of this study pertains to the fact that the same NA was used at the beginning and at the end of the year and thus familiarity effects could have affected the results. However, the amount of time that elapsed – 8 months - from the beginning to the end of the year and the fact that students did not significantly improve their scores from one point in time to the next strongly suggest that the use of the same test did not influence the results. In fact, this lack of improvement in the NA and the relatively high average attainment (15 out of 19 points) may be an indication that students had already mastered the low skills it assessed at the beginning of the school year. Also, although it is well documented that listening comprehension sets the ceiling for reading comprehension (Cutting & Scarborough, 2006; Stahl, 2004), this study did not assess listening ability due to time constraints and to classroom logistics.

Finally, although translation equivalency issues and text familiarity effects could have influenced the results, we are confident this did not occur because we followed both the same translation procedures used in the PIRLS studies and the

same criteria for text selection. Notably, a recent study indicates that the text source, which in international reading studies tends to come from an English speaking country, is not a source of bias (Grisay, Gonzalez & Monseur, 2009). For the NA, although we chose a narrative by a well-known Portuguese author, which is considered a high interest text for children, we made sure that the passage was not included in the children's language arts textbook. We believe this comparative analysis between a national and an international assessment can provide some insight into how PIRLS participating countries might evaluate the performance of their students. In particular, it can offer a way of looking at readability levels and at the emphasis national tests place on the different reading comprehension processes when compared to the PIRLS framework. Such efforts can help guide educational policy, namely in what concerns educational evaluation and related curriculum design.

ACKNOWLEDGEMENT

This research was supported by Fundação para a Ciência a e Tecnologia, Portugal.

REFERENCES

Andreassen, R. & Braten, I. (2009). Examining the prediction of reading comprehension on different multiple-choice tests. *Journal of Research in Reading, 33*, 263-283. http://dx.doi.org/10.1111/j.1467-9817.2009.01413.x

Adams, M. J. (2009). The challenge of advanced texts: The interdependence of reading and learning. In E. H. Hiebert (Ed.), *Reading more, reading better* (163-189). NY: Guilford Press.

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P.E. (2001). Comprehension skill, inference making ability and their relation to knowledge. *Memory and Cognition, 29,* 850-859. http://dx.doi.org/10.3758/BF03196414

Cain, K., & Oakhill, J. V. (1999). Inferencing making ability and its relation to comprehension failure in young children. *Reading and Writing: An Interdisciplinary Journal, 11* (5), 489-503. http://dx.doi.org/10.1023/A:1008084120205

Cardoso-Martins, C. (1995). Sensitivity to rhymes, syllables, and phonemes in literacy acquisition in Portuguese. *Reading Research Quarterly, 30(4),* 808-828. http://dx.doi.org/10.2307/748199

Catts, H. W. (2009). The narrow view of reading promotes a broad view of comprehension. *Language, Speech, and Hearing Services in Schools, 40,* 178-183. http://dx.doi.org/10.1044/0161-1461(2008/08-0035)

Catts, H. & Hogan, T. (2003). Language basis of reading disabilities and implications for early identification and remediation. *Reading Psychology, 24,* 223-246. http://dx.doi.org/10.1080/02702710390227314

Chall, J.S. (1996). Learning to read: The great debate (revised). New York: McGraw-Hill.

Chall, J., & Jacobs, V. (2003). The classic study on poor children's fourth grade slump. *American Educator*, *27*(1), 14-44.

Correia, R. (2010). Comparação dos níveis de desempenho nas Provas de Aferição [A Comparison of attainnement levels in the National Assesssments] . In F. L. Viana, R. Ramos, E. Coquet & M. Martins (Eds.), *Investigação em leitura, literatura infantil & ilustração*. *Actas do 7.º Encontro Nacional / 5.º Internacional* [CD-ROM] (pp. 139-154). Braga: Universidade do Minho.

Cutting, L. & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehen-

sion is measured. *Scientific Studies of Reading, 10*(3), 277-299. http://dx.doi.org/10.1207/s1532799xssr1003_5

Elley, W. (1992). *How in the world do students read?* The International Association for the Evaluation of Educational Achievement. Hamburg: Grindeldruck.

Ellis, N. C., Natsume, M., Stavropoulou, K., Hoxhallari, L., Van Dall, V., Polyzoe, N., Tsipa, M.L., & Petalis, M. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly, 39(*4), 438-468. http://dx.doi.org/10.1598/RRQ.39.4.5

Fernández, J. (1959). Medidas sencillas de lecturabilidad [Measures of Readability]. *Consigna, 214*, 29-32.

Foorman, B. R. (2009). Text difficulty in reading assessment. In E. H. Hiebert (Ed.), *Reading more, reading better* (231-250). NY: Guilford Press.

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

Giasson, J. (2000). *A compreensão na leitura*. [Reading comprehension]. Lisboa: Edições ASA.Gijel, M., van Bon, W., Bosman, A. (2004). Assessing reading skills by means of paper-and-pencil lexical decision: Issues of reliability, repetition, and word-pseudoword ratio**.** *Reading and Writing, 17***,** 517-536.

Gomes, I., & Castro, S. L. (2003). Porlex, a lexical database in European Portuguese.

Psychologica, 32, 91-108. Porlex: Retrived 10 October 2010 http://www.*portulex*.net Grisay, A., Gonzalez, E., Monseur, C. (2009). Equivalence of item difficulties across versions of the PIRLS and PISA reading assessments. (pp.63 – 84). IERI Monograph Series: Issues and Methodologies in Large-scale assessments, vol.2 Princeton, NJ: IEA-ETS Research Institute.

Guthrie, J. T., Wigfield, A., & Von Secker, C. (2000). Effects of integrated instruction on motivation and strategy use in reading. *Journal of Educational Psychology, 92*(2), 331-341. http://dx.doi.org/10.1037/0022-0663.92.2.331

Jenkins, R. J., Fuchs, L. S., van den Broek, Espin, C., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology, 95*, 4, 719-729. http://dx.doi.org/10.1037/0022-0663.95.4.719

Kamil, M. (2004). Vocabulary and Comprehension Instruction: Summary and implications of the National Reading Panel Findings. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (213-234). Baltimore: Paul H Brookes Publishing.

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

Mullis, I.V.S., Kennedy, A. M., Martin, M.O., & Sainsbury, M. (2006). *PIRLS 2006 Assessment framework and specification* (2nd Ed.). TIMSS & PIRLS International Study Center

Lynch School of Education, Boston College Chestnut Hill, MA: Boston College.

Mullis, I.V.S.,Martin, M.O., & Gonzalez, E.J. (2004). *International achievement in the processes of reading comprehension: Results from PIRLS 2001 in 35 Countries*. Chestnut Hill, MA: Boston College.

National Assessment Governing Board. (2007*).Reading framework for the 2009 national assessment of educational progress*. US Department of Education: Washington DC. Naggy, W. E., & Scott, J. A. (2000). Vocabulary processes. In M. L. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 269-284). Mahwah, NJ: Erlbaum.

Patel, T. K., Snowling, M., J., & de Jong, P. F. (2004). A cross-linguistic comparison of children learning to read in English and Dutch. *Journal of Educational Psychology, 96,* 785-797. http://dx.doi.org/10.1037/0022-0663.96.4.785

Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. Gough, L. Ehri, & R. Trieman (Eds.), *Reading acquisition* (pp.145-174). Mahwah, NJ: Erlbaum.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling and C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227-247). Oxford: Blackwell. http://dx.doi.org/10.1002/9780470757642.ch13

Pinnell, G. S., Pikulski, J. J., Wixson, K.K., Campbell, J. R., Gough, P. B., and Beatty, A. S. (1995). *L.istening to children read aloud.* U.S. Department of Education, National Center for Education Statistics, Washington, DC.

Pressley, M. Gaskins, I. W., & Fingeret, L. (2006). Instruction and development of reading fluency in struggling readers In S. Jay Samuels & Alan E. Farstrup (Eds.), *What Research Has to Say About Fluency Instruction* (pp. 47- 69).Newark, DE: International Reading Association.

Sainsbury, M., & Campbell, J. (2003). Developing the PIRLS reading assessment. In Martin, M. O., Mullis, I.V.S., & Kennedy, A. M. (Eds.), *PIRLS 2001 technical report* (pp. 13-27). Chestnut Hill, MA: Boston College.

Scliar-Cabral, L. (2003). *Guia prático de alfabetização* [Literacy Guide]. São Paulo: Editora Contexto.

Sénéchal, M., Ouellette, G., & Rodney, D. (2006). The misunderstood giant: On the predictive role of early vocabulary to future reading. In *Handbook of early literacy research* (pp. 173-184). NY: Guilford Press. Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundations of literacy acquisition in European orthographies. *British Journal of Psychology, 94,* 143-174.

Shiel, G. & Eivers, E. (2009). International comparisons of reading literacy: What can they tell us? *Cambridge Journal of Education, 39*(3), 345-360. http://dx.doi.org/10.1080/03057640903103736

Spear-Swerling, L. (2004). Fourth graders´ performance on a state-mandated assessment involving two different measures of reading comprehension. *Reading Psychology, 25,* 121-148. http://dx.doi.org/10.1080/02702710490435727

Stahl, S. (2004). What do we know about fluency? Findings from the national reading panel. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 187- 212). Baltimore: Brookes Publishing.

Stanovich, K. (2000). *Progress in understanding reading: scientific foundations and new frontiers*. NY: The Guilford Press.

Stanovich, K. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16*, 32-71. http://dx.doi.org/10.2307/747348

U.S. Department of Education, National Center for Education Statistics. (2003). A Content comparison of the NAEP and PIRLS fourth-grade reading assessments. NCES 2003-10, by Marilyn Binkley and Dana Kelly. Project Officer: Marilyn Binkley. Washington, DC.

U.S. Department of Education, National Center for Education Statistics. (2005). Fourth-grade students reading aloud: NAEP 2002 special study of oral reading. NCES 2006-469, Project Officer: Arnold Goldstein. Washington, DC.

Vaessen, A., Gerretsen, P., Blomert, L. (in press). Naming problems do not reflect a second, independent core deficit in dyslexia: "Double deficits" explored. *Journal of Experimental Child Psycholog.* Yuill, N., & Oakhill, J. (1991). *Children's problems in text comprehension: An experimental investigation*. Cambridge: Cambridge University Press.

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V. reis, A., Faísca, L., Saine, N., Heikki, L., Vassen, A. & Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science,* XX(X) 1-9. http://dx.doi.org/10.1177/0956797610363406

APPENDIX A – LIST OF WORDS USED IN THE WORD READING TEST

existo	exist
itinerário	itinerary
relojoeiros	locksmith
quilo	kilo
auxílio*	help
terrífico	terrifying
ágil	agile
anexo*	annex
alvoroço	rampage

proporciona	provides
veículo	vehicle
irresistível	irresistible
caixeiros*	clerks
tranquilo*	quiet
espécies	species
sustento	subsistence
recreio	playground
confessou	confessed
invejoso	envious
misterioso	mysterious

* Irregular words (word reading cannot be predicted from grapho-phonic rules)