DEVELOPING THE "LANGUAGE PROFILE TEST" FOR GREEK STUDENTS AGED 11-15 YEARS

RALIA G. THOMA

University of Thessaly

Abstract

This paper discusses the development of the "Language Profile Test," a tool for assessing the underlying component skills of reading fluency and comprehension for students aged 11–15 in the context of Greek as L1. The test consists of five subtests: three reading subtests (isolated words, pseudowords, and text), spelling (single-word spelling), comprehension, and vocabulary (cloze test). The paper aims to present the development of the test, which was based on previous research on lexical representations, specialized corpora, word-frequency lists, and cloze tests as a measure of vocabulary assessment and comprehension, reading fluency, and spelling. Special emphasis has been given to the asymmetry in the transparency of Greek orthography between the feedforward (reading) and feedback (spelling) directions that were considered for the test creation. The "Language Profile Test" was tested on a sample of 346 students. Our findings revealed that students fell into three performance categories for each subtest: high, average, and low. This classification can give teachers more insights into students' challenges regarding the underlying components of reading fluency and comprehension.

Keywords: Greek language, screening assessment, adolescent students, specialised corpora, language profile

Thoma, R.G. (2023). Developing the "Language Profile Test" for Greek students aged 11-15 years. L1-Educational Studies in Language and Literature, 22, 1-29. https://doi.org/10.21248/l1esll.2023.23.1.422

Corresponding author: Ralia G. Thoma, University of Thessaly, Department of Primary Education. Argonauton & Filellinon, Volos, 38221, Greece. email: rthoma@uth.gr

1. INTRODUCTION

It is indisputable that reading fluency is a hallmark of reading comprehension, as it helps readers make connections and fully comprehend a text (Altani et al., 2020; Clemens et al., 2017;). As Valencia et al. (2010) noted, "fluency is the ability to read text quickly, accurately, with proper phrasing and expression, thereby reflecting the ability to simultaneously decode and comprehend" (p. 271). Reading fluency and comprehension are only the tip of the iceberg, as students' proficiency in a word's identity features, namely orthography, phonology, morpho-syntax, and meaning, is strongly associated with their fluency and comprehension skills (Perfetti, 2007).
Recognizing the importance of screening assessments for the implementation of differentiated instructional models where teachers aim to meet the unique needs of each student (Durán et al., 2019), some tests for assessing one or more of the aforementioned underlying component skills of reading fluency and comprehension have been developed and validated in Greece. However, most are only commercially available and aimed at preschool or early primary school pupils.

Nevertheless, screening assessments for early and adolescent students are valuable as "linguistic difficulties affect not only academic achievement but also social relations" (Tarvainen et al., 2021, p. 2). To our knowledge, only one commercially available assessment tool, the 2nd level of the Lato test (Tzouriadou, 2008), exists for assessing students aged 8–15 literacy skills. According to the literature, however, adolescent students (aged 11–15 years old), specifically those in the upper class of primary school and junior high school (6th grade and grades 1–3 in the Greek educational system, respectively), differ from younger pupils in morphological awareness development: while awareness of inflectional morphology develops in early primary school age, awareness of derivational morphology and morphology of compounds is mainly developed during the 5th grade (10 years old) and continues to develop till the upper grades of junior high school (Carlisle, 2003; Kuo & Anderson, 2006). In addition, early and adolescent students have already developed or need to develop "higher level skills and strategies to actively engage in, read, and understand a variety of complex texts" (Swanson et al., 2016, p. 200) than their younger peers.

In the present study, we address a significant gap in applied linguistics research in the Greek language by describing the development of the "Language Profile Test". Considering Perfetti's (2007) framework of the Lexical Quality Hypothesis (hereinafter LQH), this test aims at assessing students aged 11–15 years old underlying component skills of reading fluency and comprehension.

1.1 LQH: Previous research

According to Perfetti's (2007) LQH, the quality of word representations, such as spelling, phonology, morpho-syntax, and meaning, determines comprehension (Perfetti & Hart, 2001, 2002; Perfetti, 2007). Perfetti (2007) briefly mentions that

"variation in the quality of word representations has consequences for reading skill, including comprehension" (p.357).

In high-quality representations, the main constituents of a word's identity (orthography, phonology, morphosyntax, meaning) are well specified, stable, and tightly bound together, resulting in imminent retrieval of at least one of the four features and activation of the rest of features that are associated with the same word. Readers with high-quality lexical representations free their cognitive resources from low-level processes, such as decoding, and allocate them to high-level ones, such as comprehension and writing production (Kendeou et al., 2012; Hamilton et al., 2013; Richter et al., 2013; Wolf & Katzir-Cohen, 2001). Consequently, they can read fluently, easily understand the meaning of new words, or retrieve efficiently from long-term memory the meaning of words they already know (Kendeou et al., 2012; Perfetti & Stafura, 2014). On the other hand, low lexical representations are imprecise and incomplete, resulting in a negative effect on reading fluency, comprehension, and spelling as their retrieval demands the allocation of cognitive resources to low-level processes such as decoding (Perfetti, 2007; Perfetti & Stafura, 2014).

The importance of interconnection among word constituents has been highlighted in research with adult participants. As Perfetti & Hart (2002) noted, orthography, phonology, and meaning (both vocabulary and text comprehension) are the components that differentiate skilled adult readers, while variance in orthography and phonology account more for less skilled readers. Previous studies have noted the importance of words' constituents linking vocabulary knowledge to both reading (Braze et al., 2007) and comprehension, especially when decoding is effortless (Protopapas et al., 2007[b]), while orthography has been linked to word identification processes (Andrews & Hersch, 2010).

1.2 Assessing the constituents of a word's form based on the framework of LQH

Considering the framework of LQH outlined above, we concluded that lexical quality, due to its multidimensionality, should be evaluated "through a composite of diverse factors that, whilst distinct, are interrelated" (Gómez Vera, 2016, p. 1319). For this purpose, we reviewed previous research to decide the type of measure that should be designed.

Previous research has recognized the importance of specialized corpora in language assessment (Cushing, 2017; Jang, 2014; Xi, 2017). Using materials for designing assessment tasks that are "at the right complexity level and have the desired linguistic characteristics" (Xi, 2017, p. 573), realistic examples of language usage in different contexts (McEnery & Xiao, 2010), or "reveal connections between linguistic patterning and contexts of use" (Koester, 2010, p. 67) are some of the potential advantages of basing language tests on specialized corpora.

Word frequency, which equates to the count of word occurrences in a corpus (Rayner et al., 2016; Terzopoulos et al., 2017), is one of the most important and discussed word statistics both in quantitative linguistics and in cognitive research

(Hatzigeorgiu et al., 2001; Terzopoulos et al., 2017) as it "affects processing either the word form level or conceptual level" (Vonk et al., 2019, p. 2).

Eye movements research has shown that word frequency affects the process of lexical and linguistic properties of the fixated word, with high-frequency words being skipped more often, processed, and recognized faster (shorter fixation durations) than low-frequency words (Cook & Wei, 2017; Rayner et al., 2016; Schmidtke & Kuperman, 2020). Consequently, the four main features of a high-frequency word's identity (phonology, orthography, morphology, and meaning) are easily retrieved from the mental lexicon (Vonk et al., 2019), "the mental system which includes all the information a person knows about words" (Richards & Schmidt, 2010, p. 339). This suggests that word frequency, amongst other linguistic properties, influences the organization and the links between nodes of the mental lexicon, resulting in higher network connectivity for high-frequency words (McCarthy & Miralpeix, 2020; Steyvers & Tenenbaum, 2005;), whereas low-frequency words have fewer connections (Hills et al., 2009).

The above findings validate the framework of the LQH (Perfetti, 2007), as previous research in the field of eye-tracking highlights the importance of high-quality lexical representations in a) reading (short fixation durations) (Luke et al., 2015; Taylor & Perfetti, 2016), b) orthographic precision as "successful word identification requires readers to extract the relevant features from the perceptual input and map them to existing lexical representations" (Andrews et al., 2020, p. 8), and c) the influence of lexical precision in both lexical access and reading comprehension (Andrews & Hersch, 2010; Andrews & Lo, 2012; Andrews et al., 2020; Hersch & Andrews, 2012).

Reading fluency, at the word level, is typically measured by multi-word tasks (unrelated real words or pseudowords) that are presented simultaneously, typically in columns (hereinafter word lists) (Altani et al., 2020; Protopapas et al., 2018;). Word lists are considered a valuable metric for defining a student's ability to read sequential words aloud with accuracy and speed but without a) contextual information from surrounding words (Altani et al., 2020; Wolf & Katzir-Cohen, 2001) or b) parafoveally processing of following words such as in a text format (Kliegl & Laubrock, 2018; Rayner et al., 2016), a phenomenon known as eye-voice span according to eye-tracking research (Staub & Rayner, 2007).

Lists of pseudowords, on the other hand, are considered a reading measure that gives a clear picture of letter-phoneme correspondence that the student has developed, as the latter has not formed relevant lexical representations for them, and s/he can use only the written information for decoding them (Protopapas, 2006). Therefore, visual or spelling reading is avoided, as this methodology often eliminates the vocabulary factor which can influence students' performance. This factor is considered particularly important as it ensures the validity of the measurements for older students, whose vocabulary may differ in comparison to that of younger students (Braze et al., 2007; Mouzaki & Sideridis, 2007).

Oral reading fluency should be measured using both fluency rubrics and oral reading rate if we wish to consider all the indicators of text reading fluency—rate, accuracy, and prosody (Valencia et al., 2010; Duong et al., 2011). Oral reading rate—the number of words per minute read correctly (hereinafter wcpm)—is suitable for measuring rate and accuracy (Duong et al., 2011; Valencia et al., 2010). Fluency rubrics, on the other hand, are used to measure qualitative prosody, which "provides the link between fluent oral reading and comprehension" (Valencia et al., 2010, p. 272). "The rise and falls of pitch, the rhythm and intensity placed on certain words and phrases, and the pausing between words that sometimes occurs while reading aloud" (Schwanenflugel et al., 2015, p. 10) are the main prosodic features that distinguish fluent from poor readers and result to better comprehension (Duong et al., 2011; Schwanenflugel et al., 2015; Valencia et al., 2010).

Cloze tests are defined as "any procedure whereby bits of some discourse are omitted, and the task set the examinee is to restore the missing pieces" (Oller & Jonz, 1994, p. 3) and have been used to measure reading comprehension and vocabulary at the same time. Gellert & Elbro (2013), for instance, noted that cloze tests are "sensitive to individual differences in comprehension beyond word level" (p. 25), as the student has to understand the text's ideas properly in order to fill the gaps. Similarly, cloze tests are used for quality-oriented vocabulary assessment: students have to emphasize context, understand word order relationships and semantics, and draw their attention to complex-word components, and the context that a word is commonly used in order to successfully fill the gaps (Kongsuwannakul, 2015; Reutzel & Cooter, 2011). From the above, we concluded that cloze tests could be used to assess meaning (including vocabulary meaning and text comprehension) and morphosyntax, two more constituents of the word's form according to LQH.

The simplest way to evaluate orthography is to administer dictation tasks in the form of a passage or an isolated word list (Protopapas et al., 2013[b]; Tzakosta et al., 2011). However, it is important to note that the classification scheme of errors is essential, as "errors are related to different types of information processing" (Protopapas et al., 2013[b], p. 3) and may indicate various types of difficulties. Therefore, we decided to use a single-word measure to assess students' spelling ability that would assess the orthography and morphosyntax constitutes of a word's form as mentioned in the LQH framework.

2. METHOD

2.1 Instrument

2.1.1 Creating the specialised corpus "Glossa" and the annotated Frequency-based Word List (Glossa - FWL)

The "Language profile test" comprises five subtests developed applying the corpus linguistics methodology (Biber, 2012). To develop the measures, we created the specialized corpus "Glossa", which comprises texts from the textbooks "Glossa" for grades 5 and 6 of upper primary school and "Modern Greek Language" for grades 1-3 of junior high school (nine textbooks in total), which are approved by the Greek Ministry of Education and used to teach Greek as L1 to students.

Based on Toriida (2016), we eliminated punctuation marks, foreign words, numbers (including dates), proper nouns, names of institutions or organizations, abbreviations, and special and erroneous characters (e.g., &, %) while all word tokens were converted to lowercase including proper nouns. As a result, the specialized corpus "Glossa" lists 18,383-word types and 100,680-word tokens derived from 521 texts.

It has been decided that the unit of measurement in the list will be lemmas instead of tokens, types, or word families. Lemmas are the lexical units mainly used for the development of word-frequency lists for educational purposes (Brezina & Gablasova, 2015; Milton & Treffers-Daller, 2013; Szabo, 2015; Toriida, 2016) as, according to psycholinguistic findings, "reflect the way words are stored and processed by the brain" (Milton & Treffers-Daller, 2013, p. 4).

The Natural Language Processing (NLP) service that was developed by the NLP group of the Institute for Language and Speech Processing (ILSP), namely ILSP NLP Services (Prokopidis & Papageorgiou, 2017), was used for the detection of token boundaries, the generation of parts of speech (POS), and lemma annotations for each token. The annotation resulted in 5,706 lemmas deriving from 11,904-word types, a sample size acceptable as, according to Poulos (2015), a sample's size should be at least 5,000 words to be sufficient even for mid-frequency ranges. The free corpus analysis software AntConc (version 3.5.7) (Anthony, 2018) has been used to generate the frequency lemma list "Glossa -FWL."

Consequently, the final list (Glossa—FWL) contains the 5,706 most frequent lemmas that appear in the specialized corpus "Glossa." Lemmas have been placed in descending frequency order and grouped into ten different bands according to their percentage of text coverage.

Following Nation & Coxhead (2014), we compared the lemmas of the specialized corpus "Glossa" to the lemmas of a reference corpus, namely the Educational Greek Corpus (EGC), to make sure that the specialized corpus "Glossa" doesn't contain serious omissions and has no errors. The EGC corpus comprises a general corpus, a textbook, and an educators' corpus containing more than 36,250,000 words of written texts representing modern language use . This corpus has also been awarded the

quality seals of learning objects and metadata by "Photodentro," the Greek national aggregator of educational content.

However, it is important to note that when designing educational materials, we should consider both lemma frequency and distribution within a collection of texts to accomplish greater text coverage (Pérez & Rizzo, 2014). To state this differently, keywords should be identified, “words that are unusually frequent in a small corpus compared to their normal frequency in the language” (Koester, 2010, p. 77). Nevertheless, before statistically defining keyness value, that is, before making the comparison of the "Glossa" specialized corpus to the reference corpora EGC, we used AntConc's "Concordance Plot" tool to find the lemmas that were common across all the students' textbooks that comprised the "Glossa" specialized corpus. We concluded with this decision considering that the vocabulary used in the upper grades of junior high school textbooks would be difficult for younger students, so finding the common lemmas would give extra external validity to the language test we aimed to create. The total common lemmas out of the three frequency zones and the ten different bands of text coverage were 2,348.

To statistically define the keyness value, we used the value of log-likelihood or Chi-square statistics and the significance (p-value). Based on Anthony & Gladkov (2007), the first one "provides an indicator of the keyword's importance as a content descriptor for the appeal”, while the second “represents the probability that this keyness is accidental" (p.138). The free online corpus software UCREL Significance Test System and UCREL Log-likelihood and effect size calculator of the University Centre for Computer Corpus Research on Language (UCREL) (Lancaster University, n.d.) were used to compare the frequency of Glossa-FWL words to the frequencies of the reference corpus EGC.

Common lemmas were placed in descending order according to their keyness value. The mean keyness value was also calculated, as in the language tests, we wanted to include both positive (significantly frequent items) and negative keywords (infrequent items) (Pérez & Rizzo, 2012) out of the ten bands of text coverage.

2.1.2 Developing the subtest “Word List Reading Measure”

The characteristics of the Greek orthographic transparency in the feedforward direction - from orthography to phonology - were taken into consideration for selecting the lemmas that would comprise the "Glossa" word list reading measure (Protopappas & Vlachou, 2009, pp. 994–995): "words that can be read correctly on the basis of letter sequence alone" and words that fulfill at least one of the following criteria: "phonemes spelled with two or more letters, such as <μπ> - [b]… single letters pronounced as two phonemes, such as <ξ> - [ks]… context-depend transcriptions", such as letters pronounced differently depending on adjacent letters such as <γ> in the word <άγχος> - /ˈaŋxos/ against <γ> in the word <γάτα> - /ˈɣata/… "the general phenomenon of CiV - that is, the occurrence of an [i] grapheme preceded by a consonant followed by a vowel... " such as in the word <ήλιο> (helium) pronounced [ˈilio] and

<άδεια> (permission) pronounced /ˈaðia/" and lastly homographs—words with the same spelling but different meaning, such as "<άδεια> (permission) pronounced /ˈaðia/ and <άδεια> (empty) pronounced /ˈaðja/".

Based on the above-mentioned criteria, we selected nine lemmas from each of the ten bands of text coverage. Furthermore, since Greek is a phonologically opaque language in reading (Porpodas, 2006), most of the selected lemmas were negative keywords to increase the test's difficulty. Word length was another linguistic characteristic that we considered when selecting lemmas. Eye movement research suggests that word length affects words' recognition, resulting in longer fixations and refixations for longer words, measures that reflect difficulties in word recognition (Conklin et al., 2018; Rayner et al., 2016; Tiffin-Richards & Schroeder, 2015). Consequently, word length interacts with word frequency, as long, infrequent words present longer fixations and more refixations for children (Clifton et al., 2016; Rayner et al., 2016; Tiffin-Richards & Schroeder, 2015). Therefore, the online tool Num Tool developed by Protopapas et al. (2012) was used to calculate the number of letters and phonemes for each selected lemma.

Consequently, the "Glossa" word list reading measure consisted of 90 lemmas (verbs and nouns) from 1 to 6 syllables grouped in four double-spaced, left-aligned columns (2 columns X 22 lemmas and 2 columns X 23 lemmas) of increasing difficulty and was printed on a single sheet of A4 paper, in Verdana 12pt. High-frequency lemmas (text coverage 10%–20%) were placed in the first column, while the following three columns were covered only by mid- or low-frequency lemmas that consisted of more than seven letters (text coverage 30%–90%). Table 1 below demonstrates the total number of each part of speech and the number of syllables for the total of 90 lemmas that comprised the "Glossa" word list reading measure.

Table 1. Parts of speech that comprise "Glossa" word reading list

Parts of speech	2 syllables	3 syllables	4 syllables	5 syllables	6 syllables	Total
nouns	8	20	10	8	2	**48**
adjectives	2	7	6	5	1	**21**
verbs	4	5	2	2	0	**13**
participles	0	0	1	2	0	**3**
adverbs	0	0	1	0	0	**1**
pronouns	4	0	0	0	0	**4**
Total	**18**	**32**	**20**	**17**	**3**	**90**

2.1.3 Developing the subtest "Pseudoword Reading Measure"

Based on Protopapas et al. (2006), each word of the "Glossa" word reading measure was turned into a pseudoword by "changing one or more phonemes so that the result remained phototactically acceptable" (p. 421). To address this criterion, we changed one or more consonants of each word, considering "the three main phonetic dimensions: place of articulation, manner of articulation, and voicing"

(Protopapas et al., 2006, p. 421). Furthermore, we didn't change the place of each word's lexical stress diacritic (tonos in the Greek language). As a result, we changed one consonant in two syllable long words, i.e., <πό**σ**ος> (pronounced as / pó**s**os /) to form <πότος> (/pó**t**os/), one or two consonants in three syllables long words depending on phonotactic constraints, i.e., two consonants in the word <**γ**ρα**φ**είο> (pronounced as /**ɣ**ra**f**ío/) to form <**τ**ρα**σ**είο> (/**t**ra**s**ío /), but one consonant in the word <βιαστι**κ**ός> (pronounced as /vjasti**k**ós/ to form <βιαστι**φ**ός> (/vjasti**f**ós/), two consonants on four and five syllables long words, i.e., the word <πο**λ**ιτι**κ**ός> (pronounced as /po**l**iti**k**ós/) to form <πο**κ**ιτι**μ**ός> (/po**c**iti**m**ós/).

2.1.4 Validating the subtest "Test of Spelling Achievement"

Mouzaki & Sideridis (2007; Mouzaki et al., 2008) developed a single-word measure to assess students' spelling ability (grades 2–5 in the Greek educational system, ages 7–11 years old). Sixty words that "ensured representation of key instructional units' grammar and spelling rules taught in each grade were selected from the basic vocabulary section in reading books, textbooks of previous years… and were arranged in ascending order of difficulty" (Mouzaki & Sideridis., 2007, p. 216).

Nevertheless, considering that this test was, on the one hand, developed for a younger group of students and, on the other hand, the selected words came from students' reading textbooks that are not in use anymore, we proceeded in its validation to make sure that the measure would be appropriate for the group of students aged 11–15 years.

The first stage of the validation process was to verify that all words used existed in the specialized corpus "Glossa." The verification results showed that 10 words -(<ταμείο> - /tamío/, <φιλί> - /filí/, <ματαιώνεται> - /mateónete/, <εγχειριστεί> - /ehXiristí/, <χείμαρρος> - /Xímaros/, <κυνηγητό> -/cinijitó/, <ποτίζω> - /potízo/, <μόλυνση> - /mólinsi/, <δωρεάν> - /Doreán/, <αποχαιρέτησα> - /apoXerétisa/) were not found in our corpus. To address this gap, we used the free online tool NUM Tool (Protopapas et al., 2012) to calculate the Levenshtein distance for deriving the mean orthographic distance of the 20 nearest items (orthographic neighborhood). The idea sprung from previous research (Perea, 2015; Traxler, 2012) that noted the importance of orthographic neighborhoods in activating similar units in the mental lexicon. The whole process verified that the orthographic neighbors of each word were found among the four nearest items (table 2).

Stage two of the process involved verifying that all items of the measure met the following criteria for orthographic transparency in the feedback direction—from phonology to orthography (Protopapas & Vlachou, 2009, pp. 994-995): words that: a) can be spelt correctly as each phoneme maps to one letter, or b) phonemes that are spelt differently depending on adjacent phonemes, e.g., <πλάσμα> -/plazma/, c) are identified by "morphological type for grammatical infections," d) can be spelt correctly only because we are aware of their "historical origin for word stems tracing back to ancient Greek" (diachronically systematic distinction).

Table 2. Orthographic neighborhoods of each of the 10 words of the single-word test (Mouzaki & Sideridis, 2007; Mouzaki et al., 2008).

Word of the single-word test	**Orthographic neighborhoods that exist in the specialised corpus "Glossa"**
<ταμείο> -/tamío/	1: ταμείου, ταμείον, ταμεία 2: χημείο, τομείς, **ταχείς**, ταχεία, ταμία, ταμένο, ταμείων, **σημείο**, ραφείο, λαχείο, καδμείο, βαφείο 3: ωδείο, ψυγείο, χημείου
<φιλί> - /filí/	1: φιρί, **φιλώ**, φιλμ, **φιλία**, φιλέ, φιλά, φελί 2: ψιλοί, φιλό, ψιλή, ψιλά, χιλ, χαλί, φύλο, φυλή, φυλά, φύλα, φολκ, φόλα, φις
<ματαιώνεται> - /mateónete/	2: μαχαιρώνεται, ματώνεται, ματαιώνονται, ματαιώνει, καταξιώνεται, αραιώνεται 3: χαντακώνεται, φασκιώνεται, φαντασιώνεται, ταπώνεται, περαιώνεται, παγιώνεται, μπαλώνεται, **μειώνεται**, ματώνεσαι, ματιάζεται, ματαιώσει, ματαιωνόταν, ματαιώνοντας, ματαιώνονταν
<εγχειριστεί> - /ehXiristí/	1: εχγειρισθεί 2: χειριστεί 3: χειριστείς, χειρισθεί, μεταχειριστεί, ευχαριστεί, εκνευριστεί, εγχειρίσει, διαχειριστεί 4: χωριστεί, χριστεί, χειριστών, χειριστώ, χειριστού, **χειριστής**, χειριστή, χείριστες, χειριστές, χειρισμοί, χαριστεί
<χείμαρρος> -/Xímaros/	1: χείμαρρους, χείμαρρου, χείμαρρον, χείμαρροι, χείμαρρο 2: χείμαρρό 3: χίμαιρες, χίμαιρας, χειμάρρους, χειμάρρου, **λαίμαργος** 4: χώματος, χτισίματος, χτήματος, **χρώματος**, χρίσματος, **χρήματος**, χοίρος, χίμαιρα, χιλίαρχος
<κυνηγητό> - /cinijitó/	1: κυνηγητά 2: κυνηγό, κυνηγητών, κυνηγητού, κυνηγάτε 3: υφηγητή, κυνηγών, **κυνηγώ**, κυνηγούς, κυνηγούν, κυνηγού, κυνηγός, κυνηγοί, κυνηγιού, κυνηγήσω, κυνηγήστε, κυνήγησε, κυνήγησα, κυνήγημα, κυνηγηθεί
<ποτίζω> /potízo/	2: **χτίζω**, **φωτίζω**, **τονίζω**, ποτίζει, πλουτίζω, **παίζω**, **ορίζω**, **νομίζω**, κτίζω, κομίζω, αποτίω, αποτίσω 3: ψωνίζω, ψηφίζω, **χωρίζω**, χτίσω, **χαρίζω**, φωτίσω, **φροντίζω**, φοβίσω
<μόλυνση> - /mólinsi/	1: μόλυνσης, μόλυνσή 2: όξυνση, μόλυνσής, μόλυνες, μόλυνε, μόλυναν, αμόλυντη, **άμβλυνση** 3: πόλωση, πάχυνση, όξυνσης, όξυνσή, μόχλευση, μόρφωση, μόνωη, μόνη, μόλων, μόλυβδο, μόλον
<δωρεάν> - /Doreán/	1: δωρεών, δωρεάς, **δωρεά** 2: χωράν, δωρεές 3: ωσάν, ωρών, χωρών, χωρούν, χωριών, χωρίων, χωρίον, χωριόν, χωριά, χωρεί, χωράω, χωράς, χωράνε, χωρά, χρεών
<αποχαιρέτησα> - /apoXerétisa/	1: αποχαιρέτισα, αποχαιρέτησε, αποχαιρέτησαν 2: αποχαιρέτισε, αποχαιρέτισαν, αποχαιρέτα 3: **χαιρέτησα**, αποχαιρετούσα, αποχαιρετίσω, αποχαιρετήσω,

4: χαιρέτισα, χαιρέτησε, χαιρέτησαν, αυθαιρέτησε, αυθαιρέτησαν, αποχώρησα, αποχαιρετώ, αποχαιρετούσε, αποχαιρετούσαν, αποχαιρετούν

2.1.5 Developing the subtests "Cloze Test and Oral Reading Fluency Text"

Two public informational texts (readability index of 48.51) (easy text) and 50.9, respectively (difficult text), were selected as those text genres are commonly used for assessing reading comprehension (Delgado et al., 2018). In addition, those texts were measured for their readability for upper primary and junior high school students (Giagkou, 2009), following the cloze procedure, which did not exist in the above-mentioned students' textbooks. A third text, a narrative one (readability index 41.57), also checked for its readability by Giagkou (2009), has been used for assessing students' oral fluency. This text was administered individually.

Then, the cloze tests were created, following the cloze procedure method as mentioned by Giagkou (2009):

- Adding words or phrases to have texts of 180-212 words that require 35 gaps as "an acceptable standard error of 3.3 requires at least 100 students and 35 gaps per text" (Giagkou, 2009, p. 228). The "easy" cloze test (readability index 48.51) consisted of 198 words (112 content words and 76 function words), while the "difficult" cloze test (readability index 50.9) consisted of 206 words (113 content words and 93 function words). Previous research has also highlighted that word frequency is strongly related to the difficulty level of a text (Chen & Meuers, 2016; Leroy & Kauchack, 2014). Based on those assumptions, we searched in the specialized corpus "Glossa" for the frequencies of the above-mentioned content words of both texts. Concerning the "easy" text, there were 94 out of 112 content words in the corpus (47 words (41.9%) of high and equal words of medium-frequency), while, in the "difficult text," we found 90 out of 113 content words (42 high-frequency words (46.7%), 48 medium-frequency words (53.3%)).
- Deleting every 5th word of the text, leaving the first and the last sentence of each text intact, proper names, numbers, adjectives relating to frequency and duration, and abbreviations. In those cases, we deleted the following word.

2.2 Pilot

A small-scale pilot administered the initial version of each sub-test to 117 primary and junior high school students (aged 11–15 years). Minor changes were made to some of the sub-tests due to the findings. The tests were administered individually for the "Glossa" word list measure, the pseudoword reading measure and the oral reading fluency measure. According to Seymour (2005), we should consider reading accuracy and speed in regular orthographies like Greek. So, for calculating the average time needed for reading the word list, each student's reading was recorded using a smartphone's pre-installed voice recording application.

According to our data analysis, the time limit has been set to one minute. Also we took the last 23 words off for the pseudoword reading test, as no student could read them within the one minute time limit. Moreover, 12 words that the students easily recognized while reading them were discarded. These word exceptions were necessary to address a normalization of the target group. The sub-tests of spelling achievement and the cloze tests were administered in group sessions. No revisions were needed for the oral reading fluency and spelling achievement measures. Regarding the cloze test, piloting helped us decide on the synonyms we would accept as correct answers. Following Kleijn et al. (2019), we opted for semantic scoring of cloze tests in contrast to exact scoring, as when measuring both readability and comprehension, "it seems illogical to fault a reader for filling in an acceptable answer (e.g., a synonym), rather than the original word" (p. 560).

2.3 Participants

2.3.1 Participants and setting

Participants were 346 students (196 girls and 150 boys), ages 11-15 years old (mean age 12.89, SD= 1.07 years) in the 6th grade of primary school (n=108), 1st (n=121), 2nd (n=76) and 3rd (n=41) of junior high school of ten primary schools (nine public and one private) and seven high schools (six public and one private) in Attica area. School selection followed a stratified randomization approach to include schools from higher (four primary and four junior high schools) and lower (six primary and three junior high schools) socio-economic regions in Attica. Students were selected randomly from each class, but only those whose parents gave written consent to participate in the research were included in the study.

All participating students were fluent speakers of the Greek language, had never been retained in the same grade, and did not suffer from any mental or emotional impairment that prohibited their enrollment in the regular education class of their school. Students were tested individually in the "Glossa" reading word, pseudoword measures, and oral reading fluency measures, while they were tested in groups of 20 in the cloze test and the revised measure of single-word spelling. The above-mentioned measures were administered in one or more days, depending on the number of students from each school. If a student was absent on the second day of testing, his/her test data from the previous session were excluded from the study. Data collection lasted approximately two months (April-June 2018).

The procedure was the same as in the pilot: students were tested individually, and their reading was recorded using the pre-installed voice recording application of a smartphone on word reading accuracy and speed, pseudoword reading fluency, and oral reading fluency measures. The spelling, vocabulary, and silent reading comprehension measures were administered in group sessions. Measures were administered in one or two sessions depending on how many students from each school participated in the research.

2.4 Measures

Word reading accuracy and speed were tested with the "Glossa" word reading measure. Students were instructed to name within one minute each word (90 words in total) of the printed sheet as fast as possible without making errors, starting from the top of the list of each column and moving to the bottom. Only the responses that were given within the one minute time limit were taken into consideration and were scored as follows: one point for each phonologically correct and accurate use of stress (maximum of points = 90); one point for each phonologically correct and accurate use of stress of mid or low-frequency words of 7-13 letters (maximum of points = 48). All other responses received no points, e.g., phonologically correct but inaccurate use of stress, syllables' or phonemes' omission. The maximum number of points for this measure was 138.

Pseudoword reading fluency was tested with the "Glossa" pseudoword reading measure. The instructions for students were the same as those in the "Glossa" word reading measure, but here students had to read 55 pseudowords within one minute. Only the responses that were given within the one minute time limit were taken into consideration and were scored as follows: one point for each phonologically correct and accurate use of stress (maximum of points = 55); one point for each phonologically correct and accurate use of stress in mid- or low-frequency words of 7-13 words (maximum of points = 20). All other responses received no points, e.g., phonologically correct but inaccurate use of stress, syllables' or phonemes' omission. The maximum of points for this measure was 75.

Oral reading fluency was tested with the prementioned narrative text. Students were advised to read the text orally, as they would in class. If they came to a word they could not read, they should try to do their best or skip it. No assistance was provided for unknown words. The rate was defined as the number of words students read correctly (including the stress diacritics) within one minute (wcpm). Prosody was scored—scores ranged from one to four—using the qualitative fluency rubric of Fountas & Pinell (2009), titled "Six Dimensions Fluency Rubric." This rubric comprises five dimensions that rate reading prosody and precisely rate, phrasing, intonation, pausing, stress, and the integration of the five dimensions indicated above.

To control for possible variation in scoring and following Sarris & Dimakos (2015, p. 52), we asked three different and experienced Greek primary school teachers that had no connection with this study to score all students' recordings. Then the mean value of the three scorers for each dimension was calculated. The average value of all dimensions was set as each student's prosody score. The maximum number of average points for this measure was four.

Spelling was tested with the revised single-word test (Mouzaki & Sideridis, 2007), following the guidelines of its developers as mentioned above but changing its scoring. Specifically, the group of students was instructed to write on the A4 sheet that was distributed to them (two numbered columns, 30 lines per column, with the appropriate space for writing the word). Following Mouzakis & Sideridis (2007), "during

administration, the examiner first pronounced each word in isolation and then in context to demonstrate its use. After repeating the word in isolation for the second time, the examiner asked the students to write the word in the numbered space corresponding to each word" (p. 216). We scored the test as follows: one point for each word with accurate spelling (maximum 60 points), one point for each accurate morphological spelling (maximum 48 points), one point for each accurate etymological and historical spelling (maximum 134 points), and one point for each accurate spelling of mid- or low-frequency words (maximum 33 points). The maximum number of points for this measure was 275.

Silent reading comprehension and vocabulary were tested with the two cloze tests mentioned above. Following the procedure of the pre-tests, we opted for semantic scoring. In addition, if a student had not filled in 1/3 of the gaps in one cloze test, we excluded all his/her data from the abovementioned tests. We scored the test as follows: one point for each correct word in the "easy" cloze text (maximum 35 points), one point for each correct word in the "difficult" cloze text (maximum 35 points). We also gave a bonus of five points to each student who scored higher in the "difficult" text than in the "easy" text. So, the maximum number of points a student could score from both cloze texts plus bonus points was 75 points. A summary of the maximum points per measure is shown in Table 3 below.

Table 3. Language Profile's measures and maximum points per measure

Measure	**Max. points**
"Glossa" reading measure for testing reading accuracy and speed	138
"Glossa" pseudoword reading measure for testing pseudoword reading fluency	75
Narrative test for testing oral reading fluency	4
Revised Single-word test (Mouzaki & Sideris, 2007) for testing spelling	275
Cloze tests for testing silent reading comprehension and vocabulary	75

3. RESULTS

Language profile subtests' raw and factor scores were converted to z-scores (M =0, SD = 1) using SPSS v23.0 to facilitate interpretation across the measures.

The reliability coefficient (Cronbach's alpha) for all Language Profile subtests (word and pseudoword reading, spelling, cloze test) was strong as it exceeded the 0.5 cut-off suggested by Sarafidou (2011) as a value for acceptable reliability. Concerning the subtest of oral reading fluency, both the reliability of the three scorers of the prosody scale developed by Fountas & Pinnell (2009) and each prosody's dimensions were also satisfactory ($a \geq 0.7$) (Table 4).

Table 4. Language Profile subtest and "Six Dimensions Fluency Rubric" reliability index

Language profile subtests	Cronbach's Alpha
Word reading	0.75
Pseudoword reading	0.805
Spelling	0.942
Cloze test	0.826
Six Dimensions Fluency Rubric	
Rate	0.804
Phrasing	0.806
Intonation	0.812
Pausing	0.801
Stress	0.792
Integration	0.861
Prosody score	0.977

Regarding the subtest word reading accuracy and speed (M=50, SD=10), most students (54.90%) achieved average performance values, a few students (3.76%) presented low performance, and 41.37% of students achieved a high-performance score (57.98%). Regarding lower performance compared to all other subtests, we observe a smaller value range in word reading. When looking at the average performance, the range is even smaller as many students recorded few values (Table 5).

Table 5. Students' performance in "Glossa" word reading measure subtest

Performance	Students' scores	Frequency	Percent	Cumulative Percent
	0.09	1	0.29	0.29
Low	6.36	1	0.29	0.58
	12.81	3	0.87	1.45
	19.27	4	1.16	3.76
	25.72	4	1.16	3.76
	32.17	18	5.20	8.96
Average	38.62	28	8.09	17.05
	45.08	46	13.29	30.35
	51.53	98	28.32	58.67
High	57.98	143	41.33	100.0
Total	--	346	100.0	--

In the subtest of pseudoword reading fluency (M=50, SD=10), the range in low and average performance values is relatively large. In contrast to the other subtests, only 23 students (6.65%) achieved high-performance scores. Most participants concentrated on average performance values, much more than in the "word reading test," since 306 students (88.43%) responded within this range. Table 6 below summarizes the scores for this subtest.

Table 6. Students' performance in "Glossa" pseudoword reading measure subtest

Performance	Students' scores	Frequency	Percent	Cumulative Percent
Low scores range	21.39 – 29.79	17	4.93	0.29 - 4.91
Average scores range	30.49 – 62.00	306	88.44	5-20 - 93.35
High scores	62.7	4	1.16	94.51
	63.4	19	5.49	100.0
Total	--	346	100.0	--

The results for each of the six prosody rating scale factors (Fountas & Pinell, 2009) and the overall prosody score are shown below (Table 7). Regarding the rate dimension, in 15 cases (4.33%), the scores were relatively low, while a large part of the sample achieved average (45.67%) or high performance (50%) scores. The phase dimension scores are similar to those in the rate dimension. However, the number of participants in the scoring is distributed differently. We also observed a decrease in the percentage of students who achieved high-performance scores.

Table 7. Students' performance in rate and phrase dimensions

Performance	Students' scores		Frequency		Percent		Cumulative percent	
	rate	phrase	rate	phrase	rate	phrase	rate	phrase
	16.66	18.13	2	1	0.58	0.29	0.58	0.29
Low	21.48	22.9	9	9	2.60	2.60	3.18	2.89
	26.31	27.67	4	8	1.16	2.31	4.34	5.20
	31.13	32.44	12	13	3.47	3.76	7.80	8.96
Average	35.95	37.21	22	24	6.36	6.94	14.16	15.90
	40.77	41.98	24	32	6.94	9.25	21.10	25.14
	45.60	46.75	41	45	11.85	13.01	32.95	38.15
	50.42	51.52	59	69	17.05	19.94	50.00	58.09
High	55.24	56.28	85	66	24.57	19.08	74.57	77.17
	60.06	61.05	88	79	25.43	22.83	100.0	100.0
Total	--	--	346		100.0	100.0	--	--

It is worth mentioning that in intonation, students scored the highest value (24.09) for low performance compared to other dimensions. Once more, most of the students (53.17%) achieved average-performance scores. Compared to the different dimensions, pausing had the lowest score (13.82) and the highest percentage of low-performance students (6.64%). Students' scores in both dimensions are shown in Table 8 below.

Regarding the stress dimension, students' response distribution is mainly average, while the range of average performance scores is relatively high. In the integration dimension, however, most students achieved high (48.55%) or average performance (46.54%). Table 9 below shows students' scores in both dimensions.

Table 8. Students' performance in intonation and pausing dimensions

Performance	Students' scores		Frequency		Percent		Cumulative percent	
	intonation	pausing	intonation	pausing	intonation	pausing	intonation	pausing
Low	24.09	13.82	11	2	3.18	0.58	3.18	0.58
	28.73	18.9	10	3	2.89	0.87	6.07	1.45
	--	23.98	--	4	--	1.16	--	2.60
	--	29.06	--	14	--	4.05	--	6.65
Average	33.37	34.14	9	17	2.60	4.91	8.67	11.56
	38.01	39.22	35	26	10.12	7.51	18.79	19.08
	42.65	44.3	34	39	9.83	11.27	28.61	30.35
	47.29	49.38	40	61	11.56	17.63	40.17	47.98
	51.93	--	66	--	19.08	--	59.25	--
High	56.57	54.46	61	68	17.63	19.65	76.88	67.63
	61.21	59.54	80	112	23.12	32.37	100.0	100.0
Total	--	--	346	346	100	100	--	--

Table 9. Students' performance in stress and integration dimensions

Performance	Students' scores		Frequency		Percent		Cumulative Percent	
	stress	integration	stress	integration	stress	integration	stress	integration
Low	20.4	17.74	1	2	0.29	0.58	0.29	0.58
	24.99	22.39	10	10	2.89	2.89	3.18	3.47
	29.58	27.04	10	5	2.89	1.45	6.07	4.91
Average	34.17	31.69	12	9	3.47	2.60	9.54	7.51
	38.76	36.34	41	25	11.85	7.23	21.39	14.74
	43.35	40.99	28	24	8.09	6.94	29.48	21.68
	47.94	45.64	54	44	15.61	12.72	45.09	34.39
	52.53	50.30	58	59	16.76	17.05	61.85	51.45
High	57.12	54.95	57	56	16.47	16.18	78.32	67.63
	61.71	59.60	75	112	21.68	32.37	100.0	100.0
Total	--	--	346	346	100	100	--	--

Finally, the prosody dimension (Table 10) has the broadest scores across all performances compared to the other dimensions. Most students are average performers (82.95%) and are essentially "split" into many different scores rather than many students populating a particular performance. However, the upper scores of this average range have been achieved by many students. Low percentages of students have achieved low (4.91%) or high-performance scores (12.14%).

The last parameter measured in terms of reading is students' reading rate, equal to the total number of words read correctly in one minute (Duong et al., 2011; Protopapas & Skalamboukas, 2008; Valencia et al., 2010). This last category has a relatively wide range of low, average, and high scores, but the frequency, i.e., the

number of students who achieved this performance, is small. We also observe low distribution frequencies among students, ranging from one to four for each value in this range (Table 11).

Table 10. Students' performance in prosody dimension

Performance	Students' scores	Frequency	Percent	Cumulative percent
Low scores range	16.83 – 30.20	17	4.91	0.29 - 4.91
Average scores range	30.20 – 59.46	287	82.95	5-20 – 87.86
High scores	60.29	15	4.34	92.20
	61.13	27	7.80	100.0
Total	--	346	100.0	--

Table 11. Students' performance in reading rate

Performance	Students' scores	Frequency	Percent	Cumulative Percent
Low scores range	21.47 – 30.93	17	4.91	0.29 - 4.91
Average scores range	32.33 – 64.23	313	90.46	5.20 – 95.38
High scores range	64.58 – 90.51	16	4.62	96.53 – 100.0
Total	--	346	100.0	--

Regarding the cloze test, all performances have a range of values. In addition, the low and high-performance values exceed the range observed in the abovementioned subtests. The comparable rise in the range of the high scores should be highlighted, as only a few students have achieved a specific score, but a few students have achieved a significantly extended range of scores. Consequently, the number of students with an average score evaluated increased by more than half the number of students in the score group, despite a rise in both low and high-performance values (Table 12).

Table 12. Students' performance in cloze test

Performance	Students' scores	Frequency	Percent	Cumulative Percent
Low scores range	21.86 – 34.77	17	4.91	0.29 - 4.91
Average scores range	35.76 – 59.60	296	85.55	5.78 – 90.46
High scores range	60.59 – 148.97	33	9.54	92.77 – 100.0
Total	--	346	100.0	--

Ultimately, the average and high-performance values of the spelling subtest demonstrate a relative reduction in the percentage of students within those performances. Overall, more than half of the students (84.68 %) had average performance scores, with fewer having high-performance scores (10.69%) and even fewer having low-performance scores (4.62 %) (Table 13).

Table 13. Students' performance in cloze test

Performance	Students' scores	Frequency	Percent	Cumulative Percent
Low scores range	13.02 -29.75	16	4.62	0.29 - 4.62
Average scores range	30.15 – 60.42	293	84.68	5.20 – 89.31
High scores range	60.82 – 64.41	37	10.69	90.75 – 100.0
Total	--	346	100.0	--

4. DISCUSSION

The main aim of the present study was to develop a language test that both regular and special needs teachers could use to assess students aged 11-15 years old literacy skills with Greek as L1, emphasizing the underlying component skills of reading fluency and comprehension.

Concerning word reading accuracy and speed, our results showed that the distribution of students' performance was mainly in the average performance values (54.9%), followed by the high-performance score (41.2%). In contrast, a few students (13 students, or 3.9%) had low performance. This result is consistent with previous evidence showing that reading accuracy in transparent languages such as Greek is achieved as early as in the middle elementary grades (Protopapas et al., 2007[a]). However, we would expect that most students would achieve the highest score in this subtest as previous research concerning transparent orthographies suggests an increase in the speed of reading word lists in students of upper in comparison to students at lower primary school (Altani et al., 2020).

Consequently, any differences found in students' performance, especially in average performance values, may be closely related to vocabulary knowledge (Altani et al., 2020; Protopapas et al., 2013[a]), that seem to affect the reading of words at least up to the junior high school (Protopapas, 2017).

In this regard, it is characteristic that 61.6% of the students in the sample read all the high-frequency words correctly in one minute, while only 27.5% of the students had a similar performance with the medium-frequency words. As confirmed by eye movement research, the frequency of occurrence of the word is associated with the individual's vocabulary, being a positive factor in recognizing it and consequently to student's reading skills (Clifton et al., 2016; Rayner et al., 2016; Terzopoulos et al., 2017).

Likewise, in the subtest of pseudoword reading fluency, most participants—306 in total (88.4%)—had average performance values. At the same time, a small percentage is concentrated in the low (4.9%) and high performance (6.7%). Most students did not make errors in graphophonemic correspondence but in stress assignment accuracy (87.9% of students). Specifically, only 6.6% of the students managed to correctly stress all the pseudowords consisting of 6 characters within one minute, compared to 63.6% of the students who correctly stressed the words with a corresponding number of characters within one minute in the language test "reading words." The percentages are even lower in reading pseudowords and words

consisting of 7 to 13 characters. Thus, only 5.8% of students stressed all the pseudowords correctly within one minute, compared to 27.7% who stressed the corresponding words correctly within the same time.

Our findings are in accordance with previous empirical evidence of pseudoword reading, noting that the stress error for pseudowords was proportionally disanalogous compared to word reading, even for students aged 12-15 years old (Protopapas, 2006; Protopapas et al., 2007[a]). Evidently, "stress assignment is not simply an index of general reading ability… as pseudowords do not have internal lexical representations and must be stressed by recourse to the printed diacritic" (Protopapas, 2016, p. 4). In addition, evidence from eye movement research suggests that the influence of the sublexical pathway is reflected while reading pseudowords, as "length effects are particularly large for pseudowords compared with words" (Yap & Balota, 2015, p. 31).

As previously mentioned, wcpm is largely used as a measure of rate. Nevertheless, regarding prosody, we opted for assessing the indicators of oral reading fluency separately using both wcpm and the "Six Dimensions Fluency Rubric" (Fountas & Pinell, 2009). As a result, a more detailed understanding of oral reading fluency and fluency assessment was gained. Overall, we found that the rate was strongly aligned with the separate indicators of prosody (Valencia et al., 2010). In particular, a significant portion of the sample (45%) appears to have attained intermediate values in the rate dimension, or the pace at which a reader goes through the text, with most of them (17.1%) recording a performance of 50.42. Nevertheless, the remaining half of the participants achieved relatively high scores for this category (25.4%, a score of 60.06 and 24.6%, a score of 55.24). The results of the statistical analysis are similar in both the pausing parameter and the phrase parameter, where the performance of most participants is characterized as "average" (41.3% and 52.8%, respectively) or "high" (52.1% and 41.9%, respectively). Those findings are in agreement with previous research, which highlights that in older students, as the reading rate increases, the time of reading pauses decreases (Benjamin et al., 2013; Sarris et al., 2018; Schwanenflugen et al., 2004; Schwanenflugen et al., 2015) and reading skills become automated (Protopapas & Skalamboukas, 2008). As a result, students decode with greater proficiency while utilizing larger spelling sets, e.g., at the level of syllables rather than the level of the graphemic correspondence (Sarris et al., 2018).

As expected, the results of the dimensions of intonation and stress are consistent with those of the previous dimensions. Specifically, the performance of most students can be characterized as "average" (53.2% and 55.8% of students, respectively), while "high" performance was recorded by 40.7% and 38.2%, respectively, of the participants. This is also in accordance with previous evidence showing that even students at upper junior high school have difficulties with stress assignments or in applying while reading punctuation marks (Anastasiou & Protopapas, 2015; Protopapas & Skalamboukas, 2008).

The reading rate results (wcpm) confirmed that automaticity and prosody are the two critical components of reading fluency (Benjamin et al., 2013). As expected, the

results of the reading rate were consistent with those of prosody. It was, therefore, expected that categories (high, average, and low) would also be observed in the reading rate (wcpm).

Turning now our discussion to the assessment of silent reading comprehension and vocabulary with the cloze test, there was a range of values in all performances - low, average, and high. Notably, variability in students' performance is affected by both the breadth of their vocabulary and their reading comprehension, which is directly related to their reading ability (Altani et al., 2020; Denton et al., 2011; Guo et al., 2011; Landi, 2010; Ouellette et Beers, 2010; Protopapas & Skalaboukas, 2008; Protopapas, 2016; Yildirim et al., 2014).

We would also like to account for the association between word frequency and vocabulary as predictors of word knowledge. Overall, we found that only 51. 45% of our total sample was awarded the plus five bonus points for filling in more gaps in the "difficult" cloze text, while, according to our findings, students' grade did not affect those results. As we noticed when generating the "Glossa-FWL" frequency lemma list from the "Glossa" specialized corpus, most low- and medium-frequency words are found in the junior high school textbooks for grades 1, 2, and 3. Therefore, we hypothesized that students' vocabulary would enrich as they moved through the grades. Consequently, we expected that the percentage of students who obtained the plus five bonus points would be higher in the 2nd and 3rd grades of junior high school compared to that of the other two grades. This expectation, however, was not met since only 12.43% and 5.49% of students in grades 2 and 3 of junior high school, respectively, received the bonus of five points, compared to the 17.92% of students in the 1st grade of junior high school and the 15.61% of students in the 6th grade of primary school.

From this point of view, our results are consistent with previous research that strongly associates word frequency as a metric for estimating word difficulty (Leroy & Kauchack, 2014), highlighting the influence of word frequency on reading comprehension (Chen & Meurers, 2016). Consequently, students' individual differences, with regard to the quality of their lexical representations, affect lexical precision in both lexical access and reading comprehension (Andrews & Hersch, 2010; Andrews & Lo, 2012; Andrews et al., 2020; Hersch & Andrews, 2012) as links between the nodes of the mental lexicon result in higher network connectivity for high-frequency words (McCarthy & Miralpeix, 2020; Steyvers & Tenenbaum, 2005), while low-frequency words have fewer connections (Hills et al., 2009).

Finally, performance comparisons in the spelling subtest revealed that the range of low, average, and high values still exists. However, there is a relative decrease in the frequency of individuals who have given specific performances: more than half of the participants had average performances (293 in total, 84.7%), fewer had high performance (37 people in total, 10.7%), and even fewer low performance (16 in total, 4.6%). The low percentage of students in the "low" performance is to be expected given that students of this age group have developed, although they continue to develop, the awareness of derivational morphology and morphology of

compounds (Douklias & Konstantinidou, 2010; Kuo & Anderson, 2006; Padeliadou & Rothou, 2011).

Our findings of a more significant concentration in the "average" performance were imminent due to the substantial asymmetry that characterizes the Greek language in the feedforward direction because of the historical origin of word stems tracing back to ancient Greek (Protopapas & Vlachou,2009; Protopapas et al., 2013[b]). Although students of that age group have already developed their awareness of inflectional morphology (Kuo & Anderson, 2006; Douklias & Konstantinidou, 2010), we should not disregard the spelling challenges of the historical spelling that characterizes the Greek language, especially when it is located "inside a word where students may need to recall different representations of the same phoneme" (Xanthi, 2017: 3).

Based on our data, the most striking example of the spelling challenges posed by historical spelling was found during the classification of students' individual errors: 31.5% of them did not make any grammatical (morphological) errors. The corresponding percentage for historical errors was only 1.77%. Similar conclusions regarding the number of historical spelling errors in relation to morphological errors are also mentioned in previous research (Protopapas et al., 2013[b]; Xanthi, 2017).

5. CONCLUSION

This paper described the implementation of corpus-based techniques for developing the Language profile test, which both regular and special needs teachers can use to assess students' literacy skills, focusing on the underlying component skills of reading fluency and comprehension. Despite the method's limitations (sample demographics, such as age, gender, and socioeconomic status, were not included as covariates and may have influenced the results, and no norm-referenced measures were used to assess students' cognitive abilities, such as processing speed, working memory capacity, and fluid reasoning), the analysis produced the following results:

- A specialized corpus that consists of texts that are included in the approved by the Greek Ministry of Education students' L1 textbooks "Glossa" for grades 5-6 of upper primary school and "Modern Greek Language" for grades 1-3 for junior high school. Those textbooks are used for teaching the Greek Language as L1.
- A lemmas list consisted of the 2,348 common lemmas across the above-mentioned students' textbooks.
- The Language profile test, which is addressed to students aged 11-15 years old, has been developed considering the corpus-based techniques and is freely accessible to all interested parties.

After administering the test to a representative sample of students aged 11-15, various conclusions were drawn concerning students' assessment of underlying component skills of reading fluency and comprehension as each subtest measured those. These findings emphasize the importance and effectiveness of using specialized corpora for developing language assessment tools using language material familiar to

pupils, like the one in students' textbooks that is also included in the curriculum. Likewise, as confirmed by the analysis, this practice will help teachers determine students' profiles of strengths and weaknesses in the main constituents of a word's identity (orthography, phonology, morpho-syntax, and meaning) for developing an effective intervention plan that would help them progress.

Future steps should focus on validation checks and investigating the effect that covariates, like the sample's demographics, may have with a larger sample size. Although Cormier et al. (2022) concluded that language abilities have a significant impact on cognitive test performance, we believe that future research should focus on measures such as cognitive abilities, which are linked to academic performance and predict performance on a variety of tasks, including comprehension and vocabulary (Finn et al., 2014). Considering that those measures, at least in Greece, should be administered by psychologists and not by teachers, close cooperation with experts in the field should be established. As stated in this paper, the Language Profile test has been developed mainly to be used by teachers. In this regard, the teachers' role in validating the tool —reviewing and administering the material - is considered indispensable.

ACKNOWLEDGEMENTS

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The author reports there are no competing interests to declare. The data that support the findings of this study are openly available in "figshare" (https://github.com/rallou-thoma/Language-profile-test).

REFERENCES

Altani, An., Protopapas, Ath., Katopodi, K. (2020). From Individual Word Recognition to Word List and Text Reading Fluency. *Journal of Educational Psychology, 112*(1), 22-39. https://psycnet.apa.org/doi/10.1037/edu0000359

Anastasiou, D., & Protopapas, A. (2015). Difficulties in lexical stress versus difficulties in segmental phonology among adolescents with dyslexia. *Scientific Studies of Reading, 19*(1), 31-50. https://doi.org/10.1080/10888438.2014.934452

Andrews, S., Veldre, A., & Clarke, I. E. (2020). Measuring Lexical Quality: The Role of Spelling Ability. *Behavior research methods*, *52*(6), 2257–2282. https://doi.org/10.3758/s13428-020-01387-3

Andrews, S., & Lo, S. (2012). Not all skilled readers have cracked the code: Individual differences in masked form priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 152–163. https://doi.org/10.1037/a0024953

Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of experimental psychology. General*, *139*(2), 299–318. https://doi.org/10.1037/a0018366

Anthony, L. (2018). AntConc (Version 3.5.7) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Anthony, M., Gladkov, K. (2007). Rhetorical appeals in fundraising. In D. Biber, U. Connor & A. Upton (Eds.), *Discourse on the Move: Using corpus analysis to describe discourse structure* (pp.121-151). John Benjamins Publications. https://doi.org/10.1075/scl.28.07ant

Benjamin, R.G., Schwanenflugel, P.J., Meisinger, E.B., Groff, C., Kuhn, M. R., Steimer, L. (2013). A Spectrographically Grounded Scale for Evaluating Reading Expressiveness. *Reading Research Quarterly, 48*(2), 105-133. https://psycnet.apa.org/doi/10.1002/rrq.43

Biber, D. (2012). Corpus-Based and Corpus-Driven Analyses of Languages Variation and Use. In *The Oxford Handbook of Linguistic Analysis* Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199544004.013.0008

Braze, D., Tabor, W., Shankweiler, D. P., & Mencl, W. E. (2007). Speaking up for vocabulary: reading skill differences in young adults. *Journal of learning disabilities*, *40*(3), 226–243. https://doi.org/10.1177/00222194070400030401

Brezina, V., & Gablasova, D. (2015). Is There a Core General Vocabulary? Introducing the "New General Service List". *Applied Linguistics, 36*, 1-22. https://doi.org/10.1093/applin/amt018

Carlisle, J. F. (2003). Morphology matters in learning to read: A commentary. *Reading Psychology, 24*(3-4), 291–322. https://doi.org/10.1080/02702710390227369

Chen, X., & Meurers, W.D. (2016). Characterizing Text Difficulty with Word Frequencies. In J. Tetrault, J. Burstein, C., Leacock, H., Yannakoudakis (Eds.), Proceedings of the 11th workshop on innovative use of NLP for building educational applications, 84-94. Association for Computational Linguistics.

Clemens, N. H., Simmons, D., Simmons, L. E., Wang, H., & Kwok, O. (2017). The Prevalence of Reading Fluency and Vocabulary Difficulties Among Adolescents Struggling with Reading Comprehension. *Journal of Psychoeducational Assessment*, *35*(8), 785-798. https://doi.org/10.1177/0734282916662120

Clifton, C., Jr., Ferreira, F., Henderson, J. M., Inhoff, A. W., Liversedge, S. P., Reichle, E. D., & Schotter, E. R. (2016). Eye movements in reading and information processing: Keith Rayner's 40-year legacy. *Journal of Memory and Language, 86,* 1–19. https://doi.org/10.1016/j.jml.2015.07.004

Conklin, K., Pellicer-Sanzhez, A., & Carrol, G. (2018). *Eye-Tracking : A Guide for Applied Linguistics Research.* Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108233279

Cook, A.E., Wei, W. (2017). Using Eye Movements to Study Reading Processes: Methodological Considerations. In C. Was, F. Sansosti, & B. Morris (Eds.), *Eye-Tracking Technology Applications in Educational Research*, 27-47. IGI Global. http://doi:10.4018/978-1-5225-1005-5.ch002

Cormier, D. C., Bulut, O., McGrew, K. S., & Kennedy, K. (2022). Linguistic Influences on Cognitive Test Performance: Examinee Characteristics Are More Important than Test Characteristics. *Journal of Intelligence*, *10*(1), 8. https://doi.org/10.3390/jintelligence10010008

Cushing, S.T. (2017). Corpus linguistics in language testing research. *Language Testing, 34*(4), 441-449. https://doi.org/10.1177%2F0265532217713044

Delgado, P., Vargas, C., Ackerman, R., & Salmerón, L. (2018, November 1). Don't throw away your printed books: A meta-analysis on the effects of reading media on reading comprehension. *Educational Research Review*. Elsevier Ltd. https://doi.org/10.1016/j.edurev.2018.09.003

Denton, C. A., Barth, A. E., Fletcher, J. M., Wexler, J., Vaughn, S., Cirino, P. T., Romain, M., & Francis, D. J. (2011). The Relations Among Oral and Silent Reading Fluency and Comprehension in Middle School: Implications for Identification and Instruction of Students With Reading Difficulties. *Scientific studies of reading: the official journal of the Society for the Scientific Study of Reading*, *15*(2), 109–135. https://doi.org/10.1080/10888431003623546

Douklias, S.D., & Konstantinidou, M. (2010). *Ορθογραφία: Μάθηση και Διαταραχές* [Learning and Disorders]. Gutenberg.

Duong, M., Mostow, J., Sitaram, S. (2011). Two Methods of Assessing Oral Reading Prosody. *ACM Transactions on Speech and Language Processing, 7*(4), 1-22. https://doi.org/10.1145/1998384.1998388

Finn, A. S., Kraft, M. A., West, M. R., Leonard, J. A., Bish, C. E., Martin, R. E., Sheridan, M. A., Gabrieli, C. F., & Gabrieli, J. D. (2014). Cognitive skills, student achievement tests, and schools. *Psychological science*, *25*(3), 736–744. https://doi.org/10.1177/0956797613516008

Fountas, I., & Pinnell, G. (2009). *Benchmark assessment system*. Portsmith, NH: Heinemann.

Giagkou, M. (2009). *Σώματα Κειμένων και Γλωσσική Εκπαίδευση: Δυνατότητες αξιοποίησης στη διδασκαλία της Ελληνικής και συγκρότηση παιδαγωγικά κατάλληλων σωμάτων κειμένων* [Text corpora and language education: potentials for their use in the teaching of Greek and the creation of pedagogically appropriate text corpora] [Doctoral dissertation, National & Kapodistrian University of Athens]. National Archive of PhD Thesis.

Gellert, A. S., & Elbro, C. (2013). Cloze Tests May be Quick, But Are They Dirty? Development and Preliminary Validation of a Cloze Test of Reading Comprehension. *Journal of Psychoeducational Assessment*, *31*(1), 16–28. https://doi.org/10.1177/0734282912451971

Gómez Vera, G., Sotomayor, C., Bedwell, P., Domínguez, A.M., & Jéldrez, E. (2016). Analysis of lexical quality and its relation to writing quality for 4th grade, primary school students in Chile. *Reading and Writing, 29*, 1317-1336. https://doi.org/10.1007/s11145-016-9637-9

Guo, Y., Roehrig, A. D., & Williams, R. S. (2011). The relation of morphological awareness and syntactic awareness to adults' reading comprehension: Is vocabulary knowledge a mediating variable? *Journal of Literacy Research, 43*(2), 159–183. https://doi.org/10.1177/1086296X11403086

Hamilton, S. T., Freed, E. M., & Long, D. L. (2013). Modeling Reader- and Text- Interactions During Narrative Comprehension: A Test of the Lexical Quality Hypothesis. *Discourse processes*, *50*(2), 139–163. https://doi.org/10.1080/0163853X.2012.742001

Hatzigeorgiu, N., Mikros, G., Carayannis, G. (2001). Word Length, Word Frequencies and Zipf's Law in the Greek Language, *Journal of Quantitative Linguistics, 8*(3), 175-185. https://doi.org/10.1076/jqul.8.3.175.4096

Hersch, J., & Andrews, S. (2012). Lexical quality and reading skill: Bottom-up and top-down contributions to sentence processing. *Scientific Studies of Reading, 16*(3), 240–262. https://doi.org/10.1080/10888438.2011.564244

Hills, Th., Maouene, M., Maouene, J., Sheya, A., Smith, L. (2009). Categorical structure among shared features in networks of early-learned nouns. *Cognition, 112*(3), 381-396. https://doi.org/10.1016 /j.cognition.2009.06.002.

Jang, E., Wagner, M., & Park, G. (2014). Mixed Methods Research in Language Testing and Assessment. *Annual Review of Applied Linguistics, 34*, 123-153. https://doi.org/10.1017/S0267190514000063

Kendeou, P., Papadopoulos, T. C., & Spanoudis, G. (2012). Processing demands of reading comprehension tests in young readers. *Learning and Instruction, 22*(5), 354–367. https://doi.org/10.1016/j.learninstruc.2012.02.001

Kleijn, S., Pander Maat, H., & Sanders, T. (2019). Cloze testing for comprehension assessment: The HyTeC-cloze. *Language Testing, 36*(4), 553–572. https://doi.org/10.1177/0265532219840382

Kliegl, R., & Laubrock, J. (2018). Eye-movement tracking during reading. In A. M. B. de Groot & P. Hagoort (Eds.), *Research methods in psycholinguistics and the neurobiology of language: A practical guide* (pp. 68–88). Wiley Blackwell.

Koester, A. (2010). Building Small Specialised corpora. In A. O' Keefee & M. McCarthy (Ed.), *The Routledge Handbook of Corpus Linguistics* (pp. 66-79). Routledge.

Kongsuwannakul, K. (2015). Theoretical considerations of applications and implications of concordance-based cloze tests. *Digit. Scholarsh. Humanit., 30*, 541-558. https://doi.org/10.1093/llc/fqu033

Kuo, L.-j., & Anderson, R. C. (2006). Morphological Awareness and Learning to Read: A Cross-Language Perspective. *Educational Psychologist, 41*(3), 161–180. https://doi.org/10.1207/s15326985ep4103_3

Landi, N. (2010). An examination of the relationship between reading comprehension, higher-level and lower-level reading sub-skills in adults. *Reading and Writing: An Interdisciplinary Journal, 23*(6), 701–717. https://doi.org/10.1007/s11145-009-9180-z

Lancaster University (n.d). *UCREL Significance Test System.* University Centre for Computer Corpus Research on Language. Retrieved August 14, 2023, from http://corpora.lancs.ac.uk/sigtest/#extraHelp

Lancaster University (n.d). *Log-likelihood and effect size calculator.* University Centre for Computer Corpus Research on Language. Retrieved August 14, 2023, from https://ucrel.lancs.ac.uk/llwizard.html

Leroy, G., & Kauchak, D. (2014). The effect of word familiarity on actual and perceived text difficulty. *Journal of the American Medical Informatics Association : JAMIA*, *21*(e1), e169–e172. https://doi.org/10.1136/amiajnl-2013-002172

Luke, S. G., Henderson, J. M., & Ferreira, F. (2015). Children's eye-movements during reading reflect the quality of lexical representations: An individual differences approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(6), 1675-1683 https://doi.org/10.1037/ xlm0000133

McCarthy, L., Miralpeix, I. (2020). Organizational and Formational Structures of Networks in the Mental Lexicon: A State-Of-The-Art through Systematic Review. *Languages, 5*(1). https://doi.org/10.3390/languages5010001

McEnery, T., & Xiao, R. (2010). What corpora can offer in language teaching and learning. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning (pp.*364-380). Routledge.

Milton, J., & Treffers-Daller, J. (2013). Vocabulary size revisited: the link between vocabulary size and academic achievement. *Applied Linguistics Review,* 4 (1), 151-172. https://doi.org/10.1515/applirev-2013-0007

Mouzaki, An., Sideridis, G., Protopapas, Ath., Simos, P. (2008). Διερεύνηση των ψυχομετρικών χαρακτηριστικών μιας δοκιμασίας ορθογραφικής δεξιότητας μαθητών της Β΄, Γ΄, Δ΄, και Ε΄ τάξης του δημοτικού σχολείου. [Investigation of the psychometric characteristics of a spelling test for students in grades 2, 3, 4, and 5 of primary school]. *Επιστήμες της Αγωγής, 1*, 129-146.

Mouzaki, A., & Sideridis, G. D. (2007). Poor readers' profiles among Greek students of elementary school. *Hellenic Journal of Psychology, 4*(2), 205–232.

Nation, P., & Coxhead, A. (2014). Vocabulary size research at Victoria University of Wellington, New Zealand. *Language Teaching, 47*(3), 398-403. https://doi.org/10.1017/S0261444814000111

Natural Language Processing group (2023, May 25). Natural Language Processing services developed by the NLP group of the Institute for Language and Speech Processing. http://nlp.ilsp.gr/soaplab2-axis/

Oller, J. W., & Jonz, J. (1994). Why Cloze procedure? In J. W. Oller & J. Jonz (Eds.), *Cloze and coherence* (pp. 1-20). Bucknell University Press.

Ouellette, G., & Beers, A. (2010). A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story? *Reading and Writing: An Interdisciplinary Journal, 23*(2), 189–208. https://doi.org/10.1007/s11145-008-9159-1

Padeliadou, S., Rothou, K.M. (2011). Η ανάπτυξη της μορφολογικής επίγνωσης στα Ελληνικά: μια διερευνητική μελέτη [The development of morphological awareness in Greek: an exploratory study]. In E. Kitis, N. Lavidas, N. Topintzi, T. Tsagalidis (Eds), *Selected Papers from the 19th International Symposium on Theoretical and Applied Linguistics (ISTAL 9), (pp.* 527-537). Monochromia. https://doi.org/10.26262/istal.v19i0.5529

Perea, M. (2015). Neighborhood effects in visual word recognition and reading. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 76–87). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199324576.013.7

Pérez, M.J., & Rizzo, C.R. (2014). A corpus-based proposal for the grading of vocabulary teaching materials for the legal English class. *Journal of Language Teaching and Research, 3*, 24-36.

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18*(1), 22–37. https://doi.org/10.1080/10888438.2013.827687

Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. Scientific Studies of Reading, 11, 357-383. https://doi.org/10.1080/10888430701530730

Perfetti, C.A., Hart, L. (2002). The lexical quality hypothesis. In L. Verhoven, C. Elbro, P. Reitsma (Ed.), *Studies in Written Language and Literacy, 11*(4), 189-213. https://doi.org/10.1075/swll.11.14per

Perfetti, C. A., & Hart, L. (2001). The lexical basis of comprehension skill. In D. S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 67–86). American Psychological Association. https://doi.org/10.1037/10459-004

Porpodas, C. D. (2006). Literacy acquisition in Greek: Research review of the role of phonological and cognitive factors. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 189-199). Lawrence Erlbaum Associates Inc.

Poulos, M. (2015). *Σημασιολογική Επεξεργασία της Πληροφορίας.*. [Semantic Information Processing]. Kallipos, Open Educational Publications.

Prokopidis, P., Papageorgiou, H. (2017). Universal Dependencies for Greek. *In Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017)* (pp. 102-106). Association for Computational Linguistics. https://aclanthology.org/W17-0413.pdf

Protopapas, Ath., Katopodi, K., Altani, A., & Georgiou, G.K. (2018). Word Reading Fluency as a Serial Naming Task. *Scientific Studies of Reading, 22*, 248-263. https://psycnet.apa.org/doi/10.1080/10888438.2018.1430804

Protopapas, Ath. (2017). Learning to Read Greek. In L. Verhoeven & C. Perfetti (Eds.), *Learning to Read across Languages and Writing Systems* (pp. 155-180). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316155752.001

Protopapas, Ath. (2016). From Diacritics to mental lexicon: Where is the Stress? In J. Thomson & L. Jarmuliwics (Eds.), *Linguistic Rhythm and Literacy,* (pp. 237-264). John Benjamins Publications. https://doi.org/10.1075/tilar.17.11pro

Protopapas, Ath., Mouzaki, An., Kotsolakou, A., Simos, P. (2013[a]). The role of vocabulary in the context of the simple view of reading. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 29*(2), 168–202. https://doi.org/10.1080/10573569.2013.758569

Protopapas, A., Fakou, A., Drakopoulou, S., Skaloumbakas, C., & Mouzaki, A. (2013[b]). What do spelling errors tell us? Classification and analysis of errors made by Greek schoolchildren with and without dyslexia. *Reading and Writing: An Interdisciplinary Journal, 26*(5), 615–646. https://doi.org/10.1007/s11145-012-9378-3

Protopapas, A., Tzakosta, M., Chalamandaris A. Tsiakoulis, P. (2012). IPLR : an online resource for Greek word-level and sublexical information. Lang Resources and Evaluation, *46,* 449-459. https://doi.org/10.1007/s10579-010-9130-z

Protopapas, Ath., Vlachou, E. (2009). A comparative quantitative analysis of Greek orthographic transparency. *Behavior Research Methods*, *41*(4), 991-1008. https://doi.org/10.3758/BRM.41.4.991

Protopapas, Ath., Skalamboukas, X. (2008). Η αξιολόγηση της αναγνωστικής ευχέρειας για τον εντοπισμό των μαθησιακών δυσκολιών. [Assessing reading fluency to identify learning difficulties.]. *Psychology, 15*(3), 267-289.

Protopapas, Ath., Gerakaki, S., & Alexandri, S. (2007[a]). Sources of information for stress assignment in reading Greek. *Applied Psycholinguistics, 28*(4), 695-720. https://doi.org/10.1017/S0142716407070373

Protopapas, Ath., Sideridis, G., Mouzaki, A., & Simos, P. (2007[b]). Development of lexical mediation in the relationship between reading comprehension and word reading skills in Greek. *Scientific Studies of Reading, 11*(3), 165–197. https://doi.org/10.1080/10888430701344322

Protopapas, Ath. (2006). On the use and usefulness of stress diacritics in reading Greek. *Reading and Writing: An Interdisciplinary Journal, 19*(2), 171–198. https://doi.org/10.1007/s11145-005-4023-z

Protopapas, Ath., Gerakaki, S., & Alexandri, S. (2006). Lexical and Default Stress Assignment in Reading Greek. *Journal of Research in Reading, 29*, 418-432. https://doi.org/10.1111/j.1467-9817.2006.00316.x

Rayner, K., Schotter, E. R., Masson, M. E., Potter, M. C., & Treiman, R. (2016). So Much to Read, So Little Time: How Do We Read, and Can Speed Reading Help?. *Psychological science in the public interest : a journal of the American Psychological Society*, *17*(1), 4–34. https://doi.org/10.1177/1529100615623267

Reutzel, D.R, Cooter, R.B. (2011). *Strategies for Reading Assessment and Instruction: Helping Every Child Succeed.* Pearson.

Richards, J., Schmidt, R. (2010). *Longman Dictionary of Language Teaching and Applied Linguistics* (4th ed.). Routledge. https://doi.org/10.4324/9781315833835

Richter, T., Isberner, M.-B., Naumann, J., & Neeb, Y. (2013). Lexical quality and reading comprehension in primary school children. *Scientific Studies of Reading, 17*(6), 415–434. https://doi.org/10.1080/10888438.2013.764879

Sarafidou, G.O. (2011). *Συνάρθρωση Ποσοτικών και Ποιοτικών Προσεγγίσεων. Η Εμπειρική Έρευνα* [Reconciling Quantitative and Qualitative Approaches. The Empirical Research]. Gutenberg.

Sarris, Em., Tzamakou, Ai., Markakis, J., Kaloudi, Chr. (2018). Μελέτη της αναγνωστικής ευχέρειας και της προσωδίας στην ανάγνωση κειμένου σε μαθητές Α΄ και Β΄ Δημοτικού. [A study of reading fluency and prosody in text reading in first and second grade students]. In F. Gousias (Ed.), *New Educator* (pp. 1103-1112). New Educator.

Sarris, M. & Dimakos, I. (2015). Oral reading fluency and prosody: A preliminary analysis of the Greek language. *Themes in Science and Technology Education, 8*(1), 47-62.

Schmidtke, D., & Kuperman, V. (2020). Psycholinguistic methods and tasks in morphology. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.600

Schwanenflugel, P.J., Westmoreland, M.R. & Benjamin, R.G. Reading fluency skill and the prosodic marking of linguistic focus. *Read Writ, 28***,** 9–30 (2015). https://doi.org/10.1007/s11145-013-9456-1

Schwanenflugen, P. J., Hamilton, A. M., Wisenbaker, J. M., Kuhn, M. R., & Stahl, S. A. (2004). Becoming a Fluent Reader: Reading Skill and Prosodic Features in the Oral Reading of Young Readers. *Journal of educational psychology*, *96*(1), 119–129. https://doi.org/10.1037/0022-0663.96.1.119

Seymour, P. H. K. (2005). Early Reading Development in European Orthographies. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 296–315). Blackwell Publishing. https://doi.org/10.1002/9780470757642.ch16

Staub, A., Rayner, K. (2007). Eye movements and on-line comprehension processes. In G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 327-342). Oxford Handbooks Online.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cognitive science*, *29*(1), 41–78. https://doi.org/10.1207/s15516709cog2901_3

Szabo, Cz. (2015). Introducing a Romanian Frequency List and the Romanian Vocabulary Levels Test. In: *Current Issues in Linguistic Variation: The 14th international conference of the Department of Linguistics, Vol. 2*, University of Bucharest: Bucharest University Press, pp. 303–317.

Swanson, E., Wanzek, J., McCulley, L., Stillman – Spisak, St., Vaughn, Sh., Simmons, D., Fogarty, M., Hairrell, An. (2016). Literacy and text reading in middle and high school social studies and English language art classrooms. *Reading & Writing Quarterly, 32,* 199-222. https://doi.org/10.1080/10573569.2014.910718

Tarvainen, S., Launonen, K., & Stolt, S. (2021). Oral language comprehension interventions in school-age children and adolescents with developmental language disorder: A systematic scoping review. *Autism & Developmental Language Impairments*. https://doi.org/10.1177/23969415211010423

Taylor, J. N., & Perfetti, C. A. (2016). Eye movements reveal readers' lexical quality and reading experience. *Reading and Writing: An Interdisciplinary Journal, 29*(6), 1069-1103. https://doi.org/10.1007/s11145-015-9616-6

Terzopoulos, A. R., Duncan, L. G., Wilson, M. A., Niolaki, G. Z., & Masterson, J. (2017). HelexKids: A word frequency database for Greek and Cypriot primary school children. *Behavior research methods*, *49*(1), 83–96. https://doi.org/10.3758/s13428-015-0698-5

Tiffin-Richards, S. P., & Schroeder, S. (2015). Word length and frequency effects on children's eye movements during silent reading. *Vision Research, 113*(Pt A), 33–43. https://doi.org/10.1016/j.visres.2015.05.008

Toriida, M.C., (2016). Steps for Creating a Specilaized Corpus and Developing an Annotated Frequency - Based Vocabulary List. *Tesl Canada Journal, 34*(11), 87-105.

Tzakosta, M., Christianou, Ch., Kalisperaki, F. (2011). The role of morphological perception in learning the orthograhic system of Greek: Is there anything that needs to change? Studies in Greek linguistics 31. Proceedings of the 31th Annual Meeting of the Department of Linguistics, School of Philology, Faculty of Philosophy, A.U.Th., 541-553.

Tzouriadou, M. (2008). Ψυχομετρικό κριτήριο γλωσσικής ικανότητας -επάρκειας για παιδιά και εφήβους. [Psychometric criterion of language proficiency for children and adolescents]. University of Thessaly.

Traxler, J.M. (2012). *Introduction to psycholinguistics. Understanding language science.* Wiley - Blackwell.

UCREL Significance Test System. (2023, May 25). Chi-squared test, log-likelihood test and Fischer exact test for any kind of contingency table, using R. http://corpora.lancs.ac.uk/sigtest/#extraHelp

UCREL Log-Likelihood and effect size calculator. (2023, May 25). https://ucrel.lancs.ac.uk/llwizard.html

Valencia, Sh. W., Smith, A.T., Reece, A.M., Li, M., Wixson, K.K., Newman, H. (2010). Oral Reading Fluency Assessment: Issues of Construct, Criterion, and Consequential Validity. *Reading Research Quartely, 45*(3), 270-291. https://doi.org/10.1598/RRQ.45.3.1

Vonk, J., Jonkers, R., Hubbard, H. I., Gorno-Tempini, M. L., Brickman, A. M., & Obler, L. K. (2019). Semantic and lexical features of words dissimilarly affected by non-fluent, logopenic, and semantic primary progressive aphasia. *Journal of the International Neuropsychological Society : JINS*, *25*(10), 1011–1022. https://doi.org/10.1017/S1355617719000948

Durán, LK., Wackerle-Hollman, A.K., Kohlmeier, Th.L., Brunner, St.K., Palma, CJ., Callard, Ch., H. (2019). Individual Growth and Development Indicators-Español: Innovation in the development of Spanish oral language general outcome measures. *Early Childhood Research Quarterly* 48, 155-172. https://doi.org/10.1016/j.ecresq.2019.02.001

Wolf, M., Katzir - Cohen, T. (2001). Reading Fluency and Its Intervention. *Scientific Studies of Reading, 5*(3), 211-239. https://psycnet.apa.org/doi/10.1207/S1532799XSSR0503_2

Xanthi, St. (2017). Ποιοτική ανάλυση λαθών μαθητών Δ΄ - Στ΄ δημοτικού σε ορθογραφικά έργα υπαγόρευσης και ελεύθερης γραφής. [Qualitative analysis of 4th - 6th grade students' errors in dictation and free writing spelling tasks]. *Έρευνα στην Εκπαίδευση* [Research in Education]*, 6*(1), 1-17.

Xi, X. (2017). What does corpus linguistics have to offer to language assessment? *Language Testing*. *34*(4),565-577. https://doi.org/10.1177%2F0265532217720956

Yap, M. J., & Balota, D. A. (2015). Visual word recognition. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 26–43). Oxford University Press. https://psycnet.apa.org/doi/10.1016/j.jml.2009.02.001

Yildirim, S., Can, R., Turkyilmaz, M. (2014). Relations among Contextual Silent Reading Fluency, Non-Contextual Silent Reading Fluency, and Reading Comprehension: A Path Analysis of fifth-Grade Readers. *Procedia - Social and Behavioral Sciences, 116*, 4010-4014. https://doi.org/10.1016/j.sbspro.2014.01.882