

ISRG Journal of Education, Humanities and Literature (ISRGJEHL)



ISRG PUBLISHERS

Abbreviated Key Title: ISRG J Edu Humanit Lit

ISSN: 2584-2544 (Online)

Journal homepage: <https://isrgpublishers.com/isrgjehl/>

Volume – I Issue- IV (July – August) 2024

Frequency: Bimonthly



THE RELATIONSHIP BETWEEN READABILITY AND LANGUAGE LEVEL

CHRYSOVALANTOU KAPETA (Ph.D.)^{1*}, APOSTOLOS KLONIS (Ph.D.)²

^{1,2} Department of Italian Language and Literature, Aristotle University of Thessaloniki, Greece

| **Received:** 02.08.2024 | **Accepted:** 04.08.2024 | **Published:** 07.08.2024

*Corresponding author: CHRYSOVALANTOU KAPETA (Ph.D.)

Department of Italian Language and Literature, Aristotle University of Thessaloniki, Greece

Abstract

In order to measure the Readability grade and the Language Level in written production, 316 texts from the state certificate for Italian language in Greece (KPG) were selected between May 2015 and November 2016. Specifically, from 1000 randomized KPG notebooks, a total of 80 notebooks of language level B1 and B2 were used, that were first digitized in manual form. In the second phase, these texts were measured using the READ-IT tool, and in the last phase, the statistic and factor analysis have been completed through the SPSS.24 software. The main target was to research if there is any impact on the language level according to Readability measurement, and vice versa. Briefly, Greek candidates at level B produce texts in Italian depending on specific characteristics, such as grammar, vocabulary, and syntax. The results lead to the fact that all this new data determines the Language Level and degree of difficulty in written production. Furthermore, the final results are part of an existing tool named trat.exe used by the University EKPA and Aristotle University of Thessaloniki to measure the Readability in accordance to Language Level regarding the exams of Italian language of KPG. Of positive significance would be the future deepening of the parameters of writing with the ultimate goal of developing even more advanced software like trat.exe and to achieve a global data base which would lead to more valid and fair tests for non-native speakers when being evaluated according to elements found in the present study.

Keywords: Exams, Non-native Speaker, Software, Written Production.

Introduction

Appropriate tools lead to valid and fair measurement, processing and evaluation of texts produced by foreigners in a different language (Elder & Harding, 2008: 341-342). In addition, this data is very important in order to be able to create exam instructions and tests depending on language skills (in this case B1 and B2 based on the Common European Framework of Reference for Foreign

Languages) and level of difficulty, i.e. the readability grade (Lenzner, 2014: 678-681). The discovery of such data is necessary for language levels (A1-C2) and for different languages in order to create more advanced text evaluation software (Beacco, 2017: 9-19). Furthermore, the two largest Greek universities EKPA and the University of Aristotle in Thessaloniki already use a very

important measurement tool called *trat.exe* (Klonis, 2019), with which measurements of the Readability level for the preparation of Italian language exams of the State Certificate of Language Proficiency in Greece (KPG) are proceeded. By measurement we mean the use of readability formulas that use algorithms to calculate the number of word and sentence lengths, verbs, nouns, adjectives, etc. and thus measure the level of difficulty and language level of a written text.

This newly developed software, or *trat.exe* (Klonis, 2019), is a reliable and valid digital tool for assessing the difficulty of Italian language tests. It is used to develop test topics of various levels of language proficiency certification by measuring texts depending on the language level and difficulty grade. In other words, it measures and calculates the Readability level of the texts in Italian. In particular, the web software created for this purpose has the following functions:

- The reader can enter the text and the difficulty level results will be extracted by the software.
- User can see some additional features of the system such as the relevance between number of words, syllables, characters (question or exclamation mark, period, etc.), sentences, syllables per word, and words per sentence.
- It is also possible to read texts on an entire website by entering the corresponding electronic address.
- Finally, it is possible to measure the readability of doc, docx and PDF files when these files are uploaded on the software (Klonis, 2019).

Very interesting variables have been arrived at in order to be able to develop valid (validity) and fair (fairness) tests (Elder & Harding, 2008).

Purpose Of The Study

In the late 19th and early 20th centuries, the advent of public education brought unprecedented numbers of students. Consequently, it was necessary to be able to distinguish higher education students from those destined for work camps and factories (Tierney, 2013). The high regard for scientific technique during this period supported the general assumption that standardized tests were inherently fairer than the subjective methods (i.e., oral recitation, essays, etc.) that had previously been used to evaluate student performance. However, by the 1960s, questions about test fairness were sufficiently vociferous to attract the attention of the measurement community (Tierney, 2013).

According to Hoorn and van Wijngaarden (2010), the word “quality” is often used to indicate the correctness or accuracy of information; however it is also used interchangeably with reliability. In our opinion, this indicates that quality should be broken down into a set of quality indicators. Furthermore, fairness and reliability can be highly related concepts. In this case, many aspects cited to detect reliability actually indicate correctness.

The correctness of the information comes close to the truth. The reliability of the source indicates how seriously the content should be taken by the reader. Readability, therefore, is a composite of ease and reading level (Hoorn & van Wijngaarden, 2010).

Regarding validity, Gyll and Ragland (2018) argue that the objective of assessment is to ensure that a comprehensive evidence-based learning package exists to document not only knowledge proficiency, but also reasoning patterns, skills of

performance and behavior of students seeking a particular degree.

According to Annerstedt and Larsson (2010), instead of addressing validity issues, due to the lack of clear and concise guidance, teachers say it is difficult to carry out assessments fairly. Teachers argue that they have not been adequately informed about why they should formulate assessment criteria or how to check it. They also state that very little in-service training has been given regarding evaluation. In addition, they claim that the guidance provided by national authorities has been indistinct, unclear and somewhat difficult to understand (Annerstedt & Larsson, 2010: 107).

Consequently, there are two different targets regarding the present study that must be discovered:

- I. The criteria that define the language level of a written production.
- II. 2. The criteria that define the difficulty level of a written production.

Both axes lead through the use of different methods and systems, such as readability formulas, to the collection of specific fundamental data. Furthermore, all the new information collected leads to the final product of the research, i.e. all the textual characteristics of written production with which someone wants to learn, improve or teach a foreign language.

Research Questions

Regarding the issues we often face when we read a text, they are more general but substantial:

1. How can we distinguish a text according to language levels (A1-C2)?
2. What are the readability advantages for the Italian language and how can they have been exploited for the discovery of a new tool for measuring and evaluating a text such as the READ-IT tool (Dell'Orletta, Montemagni & Venturi, 2011: 75 -76)?
3. Why is it important to find a safe and reliable method to classify a text?

The concept of the language level of a text constitutes the starting point of this research, because in the end all the results revolve around the same rules: 1. how to create a test of reasonable evaluation for non-native speakers 2. how to write a text according to the degree of difficulty and language level having at hand fundamental criteria that emerge from the present study.

Review Of Related Literature

For the organization of language learning and the public recognition of results, there is a consensus between the number and nature of language levels. In other words, the European framework for foreign languages is composed of six general levels, which should be mentioned in table 1 below (Mariani, 2014), and adequately covers the space of learning European languages.

A	B	C
Elementary level	Intermediate level	Advanced level
A1: (<i>Breakthrough</i>)	B1: (<i>Threshold</i>)	C1: (<i>Effective Operational Proficiency</i>)
A2: (<i>Waystage</i>)	B2: (<i>Vantage</i>)	C2: (<i>Mastery</i>)

Table 1: Language Levels A-C

This grid, then, can be found in the CEFR which proposes six common levels and interprets each linguistic-communicative competence separately. For each level there is a series of indicators to observe to verify the level of competence, and in almost all cases it involves "knowing how to do with the language" (Balboni, 2012: 13). The six levels of competence follow with their terminology in English (Diadori, 2003: 10-11).

- ✓ A1 = Contact level (Breakthrough)
- ✓ A2= Survival level (Waystage)
- ✓ B1= Threshold level
- ✓ B2= Level of progress (Vantage)
- ✓ C1= Effectiveness level (Proficiency)
- ✓ C2= Mastery level

Each of these levels is governed by specific principles and rules, but at the same time strictly bound to the principles and rules that are included in each level (Brugè, 2000: 42).

B1 detects the ability to "go through a text to find information", while at level B2 comes speed of reading, autonomy and the ability to retrieve more salient points. It is precisely at this last level that learners seem to arrive. A priori, an evaluation based on similar texts should not be difficult to set up (Jamet, 2010: 83).

Hyunsook, Hirvela (2004: 257-283) and Kennedy and Miceli (2010: 28-44) propose that a corpus oriented according to suitable tasks that contain vocabulary proposed by the teacher can help improve a student's written production.

The enormous concentration required to undertake these twin tasks limits the amount of time one can devote to them. Although these are different professions as mentioned before, starting from the fact that one needs 10,000 hours to reach a level of excellent knowledge and perfection, the method known as content-based language teaching involves teaching subjects such as mathematics, geography and others in foreign languages (Eaton, 2011: 6). Starting then from the fact that if we consider that fluency is the same thing as "being" an expert in speaking a language, then a student can well invest 10,000 hours in linguistic studies to achieve fluency (Eaton, 2011 : 6).

Methodology

The selection of authentic, valid and true texts was challenging because it required thinking about four characteristic factors.

1. Choosing texts already produced by exams in which a large part of examinees from all over Greece participated.
2. Using texts from recent years to discover new, more effective solutions to contemporary problems and needs.
3. Finding the technological way to digitize them without altering their content.
4. Analyzing written productions using more scientific tools such as READ-IT and SPSS 24.

The investigation is divided into three phases: the first begins with the samples of the Greek State Certification for the Italian language (KPG) exams. In particular, 316 written productions were chosen (160 of the B1 level and 156 of the B2), drawn from eighty unknown individuals.

During the second phase, probably considered the most challenging part of the research, 316 produced texts must have been digitized. Without having the possibility of using some electronic instrument (tool) of maximum security, accuracy and validity, it was preferred to write all the texts produced manually with great care and accuracy (316 in total) using Word (Windows 2010) .

During the third phase, the Gulpease index and READ-IT tool were used to process all the data of the texts produced. Through these both tools, the variables were found, with which the table was compiled with all the important variables of the fourth phase for which the IBM SPSS STATISTICS VERSION 24 software was chosen.

In the fifth and final phase, all the variables collected from the SPSS table were analyzed, using SPSS graphs and tables to arrive at the conclusions and results of the hypotheses mentioned in the research questions chapter before. For the research, we see in table 2 the most necessary data in a descriptive way:

Number of analyzed texts	316
Source	Greek State Certificate
Time frame	May 2015-November 2016
Language Level	B1, B2
Formulas/Tools used	Gulpease, READ-IT
Program of statistical analysis	SPSS.24

Table 2: Overview table of data for this research

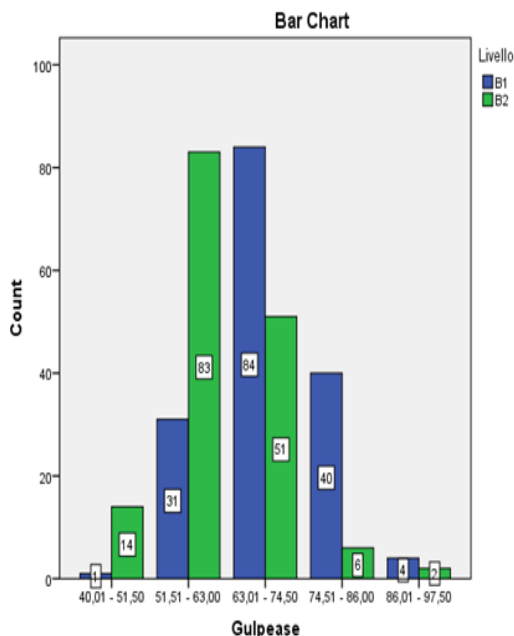
Results And Discussion

Subsequently, reference is made to the research results through which an indicative link is analyzed, i.e. the relationship between language level and degree of difficulty concerning produced texts by non-native speakers of Italian language.

Gulpease * Livello Crosstabulation

Count		Level		Total
		B1	B2	
Gulpease	40,01 - 51,50	1	14	15
	51,51 - 63,00	31	83	114
	63,01 - 74,50	84	51	135
	74,51 - 86,00	40	6	46
	86,01 - 97,50	4	2	6
Total		160	156	316

Table 3: Total result of the Gulpease index according to Language Level B1 & B2



Graph 1: Illustration of the total result of the Gulpease index according to Language level B1 & B2

The Gulpease index measures the ease of the text. Consequently, the closer it is to 100%, the easier a text is. Table 3 and graph 1 demonstrate that for the Gulpease formula, i.e. the measurement of word and sentence length for the B1 level, we find 84 written productions of B1 and 51 of B2 which make up the majority (63.01-74, 50%). As a result, many texts exceed 50% of the Gulpease index, which may show that many Greek candidates have difficulty using long words and sentences. According to this data, there are texts produced in which polysyllable words are rarely used (e.g. *importante, indimenticabile, affascinante, preoccupante, professore, proseguire, avvertire, consumatore*, etc.). The same also happens in the case of sentences which often include the simplest form, i.e. short sentences containing a subject, a verb and an adjective. This information perhaps leads to results with lower language levels and degrees of text difficulty.

Since in multivariate analysis there is an ever-expanding set of methods for data analysis that includes a wide range of possible situations, seeking new investigations, in this research, Principal Component Analysis (Basto & Pereira, 2012) and Common Factor Analysis (Principal Factor Analysis) (Floyd & Widaman, 1995) are used.

Correlation Matrix^a

	Basic Vocabulary	Fundamental vocabulary	Vocabulary of high Use	Vocabulary of high Availability	Lexical Density
Basic Vocabulary	1,000	,480	-,174	-,027	,110
Fundamental vocabulary	,480	1,000	-,711	-,238	-,009
Vocabulary of high Use	-,174	-,711	1,000	-,133	-,070
Vocabulary of high Availability	-,027	-,238	-,133	1,000	-,002
Lexical Density	,110	-,009	-,070	-,002	1,000

Sig. (1-tailed)	Lexical Density	Vocabulary of high Availability		Vocabulary of high Use		Basic Vocabulary	Fundamental vocabulary	Vocabulary of high Use	Vocabulary of high Availability	Lexical Density
		,319	,000	,001	,000					
,025	,434	,106	,486	,009	,009	,001	,000	,009	,009	,070
,434	,106	,486	,009	,009	,009	,001	,000	,009	,009	-,002
,106	,486	,009	,009	,009	,009	,001	,000	,009	,009	-,002
,486	,009	,009	,009	,009	,009	,001	,000	,009	,009	-,002
,009	,009	,009	,009	,009	,009	,001	,000	,009	,009	-,002
,009	,009	,009	,009	,009	,009	,001	,000	,009	,009	-,002
,009	,009	,009	,009	,009	,009	,001	,000	,009	,009	-,002
,009	,009	,009	,009	,009	,009	,001	,000	,009	,009	-,002

a. Determinant = ,244

Table 4: Matrix correlation between the variables “Basic Vocabulary”, “Fundamental vocabulary”, “Vocabulary of High Use”, “Vocabulary of high Availability” and “Lexical density”

The component table contains values only for the three relevant factors. Then, these values are also known as factor loadings.

On table 4 (Correlation Matrix) of the factor analysis, five variables from the lexical sector of the SPSS table are compared. The deciding factor must have a value greater than 0.0001 to lead to a correlation between the variables. In this case, the value is ,244, i.e., 24%, showing that the variables are probably correlated. The highest correlation values occur between basic vocabulary and vocabulary of high use (-174). The higher the percentage of basic vocabulary is, the lower the percentage of vocabulary of high use, and vice versa. The same also happens in the case of fundamental vocabulary and vocabulary of high use (-,711), and then, in the case of fundamental vocabulary and vocabulary of high availability (-238).

However, the significance value, when it is 0.000, then shows a very high correlation, as in the case of basic vocabulary, fundamental vocabulary, vocabulary of high use and high-availability vocabulary. Only the lexical density seems not to be of great significance (>0.000).

The negative sign (-) shows that there is probably an inversely proportional correlation between two variables. The more vocabulary of high use increases, the less basic vocabulary appears to be, and vice versa. The more fundamental vocabulary increases, the less vocabulary of high availability appears to be, and vice versa.

versa. The more fundamental vocabulary increases, the less vocabulary of high use appears to be, and vice versa. Lexical density is the variable that does not seem to influence any of the other four elements, because it is not included in almost any important correlation with other variables.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.341
Bartlett's Test of Sphericity	Approx. Chi-Square	441,189
	df	10
	Sig.	.000

Table 5: KMO test of the variables “Basic Vocabulary”, “Fundamental vocabulary”, “Vocabulary of high use”, “Vocabulary of high Availability” and “Lexical density”

This is followed by the KMO and Bartlett's Test table (Table 5), where the value for KMO should normally be more than 0.5 to be important. The higher this value is, the better for correlation reasons. Instead, the importance value (significance value) must be 0.000, if variables correlated with each other are sought, as in this search.

Conclusions

After examining all these characteristics, we can make a differentiation between the levels of linguistic competence in the texts produced by Greek users of the Italian language with respect to grammatical, lexical and syntactic factors or typical criteria, found throughout the research and especially those presented in the tables and the graph mentioned above. These characteristics demonstrate their influence on a text produced so that it can be differentiated according to level B1 or B2.

The variables that concern the lexical elements (“Basic vocabulary”, “Fundamental vocabulary”, “Vocabulary of High Use” and “Vocabulary of High Availability”) are very important for our investigation, since they increase the Language level and the text difficulty. These are variables with very high values according to which it turns out that in the 316 written productions of this research the lack of less frequent vocabulary can be considered expected, so that the composers are Greek, in short their mother tongue is not Italian.

By realizing this consideration, with the greater and more correct use of these factors, texts of higher levels could be produced, in our case texts of the intermediate level. According to this hypothesis, we can justify that many Greek users have difficulty using lesser-

known words, not having Italian as their mother tongue and, in many cases, not even as their first foreign language. In this respect, texts of lower readability level are often produced which subsequently leads to less satisfactory results.

A tool called *trat.exe*, similar to READ-IT, for which the present search is also partly performed, was created by researchers from the Aristotle University in collaboration with the Italian Department of the EKPA University to develop, measure and control the written tests of the KPG exams (Klonis, 2019). Finally, we hope to spread it both in Greece and abroad, also expanding it into different languages and language levels.

In particular, the online software created for this purpose has the following functions:

- the reader can enter his text and the software will export to him results and scores from different types of readability and also the level of difficulty regarding the tests;
- user can see some additional system functions such as word counters, syllables, characters, sentences, characters per word, syllables per word, words per sentence;
- there is also the possibility of reading texts as a whole on the website, when the user types the corresponding electronic address;
- and finally, it is possible to measure the readability of .doc, .docx, and .pdf files, if these files are published on the software (Klonis, 2019: 1239-1240).

This online software is called *trat.exe* (or better “text readability analysis tool”) and we hope that it will also be developed for other languages such as German, English, Spanish, etc. Furthermore, it will be very interesting and essential to measure and process tests using this electronic tool also for all language levels (A, B and C).

Future scientists are invited to search also data from other countries or even worldwide in order to create a global data base that might help researchers or teachers to construct more valid and fair tests for non-speakers.

Acknowledgements

None

References

1. Annerstedt, C., Larsson, S., 2010. ‘I have my own picture of what the demands are ...’: Grading in Swedish PEH – problems of validity, comparability and fairness. In *European Physical Education Review*, 16 (2), 97-98.
2. Balboni, P. E., 2012. *Le sfide di Babele-Insegnare le lingue nelle società complesse*, UTET Università (forth edition).
3. Basto, M., Pereira, J. M., 2012. An SPSS R-Menu for Ordinal Factor Analysis. In *Journal of Statistical Software*, 46(4), 1-29.
4. Beacco, J.-C., 2017. Il Quadro Comune Europeo di Riferimento per le Lingue-Un manuale-Contributo all'insegnamento delle Lingue e al Piano Trentino Trilingue. In *Trento: IPRASE-Istituto Provinciale per la Ricerca e la Sperimentazione Educativa*, Working Paper (5).

5. Brugè, L., 2000. *Teoria linguistica e insegnamento della grammatica*. Roma: Bonacci Editore.
6. Dell'Orletta, F., Montemagni, S., Venturi, G., 2011. READ-IT: Assessing Readability of Italian Texts with a View to Text Simplification. In *Association for Computational Linguistics*, 73-83.
7. Diadori, P., 2003. *Orientamento linguistico degli immigranti in Italia*. Retrieved 6/2/2020, from shorturl.at/amnL0.
8. Eaton, S. E., 2011. How long does it take to learn a second language? Applying the "10.000-hour rule" as a model of fluency. In *English Language Teaching*, 11(2).
9. Elder, C., Harding, L., 2008. Language testing and English as an international language. In *Australian Journal of Applied Linguistics*, 31 (3), 341-343.
10. Floyd, F. J., Widaman, K. F., 1995. Factor Analysis in the Development and Refinement of Clinical Assessment Instruments. In *American Psychological Association*, 7(3), 286-299.
11. Gyll, S. Ragland, S., 2018. Improving the validity of objective assessment in higher education: Steps for building a best-in-class competency-based assessment program. In *The journal of competency-based education*, 3(1), 1-8.
12. Hoorn, J. F., van Wijngaarden, T. D., 2010. Web Intelligence for the Assessment of Information Quality: Credibility, Correctness, and Readability. In *IntechOpen*, 306-308.
13. Hyunsook, Y., Hirvela, A., (2004). ESL student attitudes toward corpus use in L2 writing. In *Journal of Second Language Writing* 13.
14. Jamet, M.-C. (2010). Intercomprensione, Quadro comune europeo di riferimento per le lingue, Quadro di riferimento per gli approcci plurilingui e valutazione. In *Synergies Europe*, 5, 80.
15. Kennedy, C., Miceli, T., 2010. Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. In *Language Learning & Technology*, 14(1).
16. Klonis, A., [Κλώνης, Α.], 2019. Κατασκευή λογισμικού επιπέδου γλωσσομάθειας και μέτρησης δυσκολίας τεστ. In *Πρακτικά Εργασιών 5ου Διεθνούς Συνεδρίου, για την Προώθηση της Εκπαιδευτικής Καινοτομίας*, Λάρισα 11-13 Οκτωβρίου 2019, 1239-1245. Retrieved 16/12/2020, from shorturl.at/eAFO3.
17. Mariani, L., 2014. *Il Quadro Comune Europeo di Riferimento e la sua valenza formativa*. Retrieved 27/3/2020, from shorturl.at/nqIR6.
18. Lenzner, T., 2014. Are Readability Formulas Valid Tools for Assessing Survey Question Difficulty? In *Sociological Methods & Research*, 43(4), 678-681.
19. Tierney, R. D., 2013. Fairness in classroom assessment. In *J. H. McMillan (Ed.). SAGE Handbook of Research on Classroom Assessment*. Thousand Oaks, CA: Sage Publications, 125-144.