UNIVERSIDADE DE SÃO PAULO

READABILITY FORMULAS APPLIED TO TEXTBOOKS IN BRAZILIAN PORTUGUESE

TERESA B.F. MARTINS
CLAUDETE M. GHIRALDELO
M.GRAÇAS V. NUNES
O.N. OLIVEIRA JR.

N. 28

NOTAS

Instituto de Ciências Matemáticas de São Carlos

Instituto de Ciências Matemáticas de São Carlos

ISSN - 0103-2577

READABILITY FORMULAS APPLIED TO TEXTBOOKS IN BRAZILIAN PORTUGUESE

TERESA B.F. MARTINS
CLAUDETE M. GHIRALDELO
M.GRAÇAS V. NUNES
O.N. OLIVEIRA JR.

N. 28

NOTAS DO ICMSC
Série Computação

São Carlos
Jun./1996

Readability Formulas Applied to Textbooks in Brazilian Portuguese

Teresa B.F. Martins[1], Claudete M. Ghiraldelo[2], M. Graças V. Nunes[3] and O.N. Oliveira Jr.[2]

[1]Departamento de Letras, Unesp, Araraquara, SP. - Brazil

[2]Instituto de Física de São Carlos, USP, São Carlos, SP - Brazil

[3]Departamento de Computação e Estatística, ICMSC, USP, São Carlos, SP - Brazil

Abstract

We report on the first attempts of applying readability formulas to texts in Brazilian Portuguese. A software tool was developed which permitted one to obtain readability Flesch scores from a random selection of passages from textbooks which covered from primary school to introductory texts for college students. The scores were shown to decrease for progressing school grades, this decrease being sharper in those texts directed at the first years of primary school. These scores were practically leveled off when texts from upper high school and first years of college were compared. In order to compare with the scores ranges usually obtained for texts in English, a shift of 42 points was made in the Flesch scores for the Portuguese texts. The reason for this is that words in Portuguese contain in average a higher number of syllables than English words, which artificially deflates the Flesch scores. Other similar formulas, such as the ARI, Kincaid and Coleman-Liau, were also employed in subsidiary tests, the results leading to the same conclusions drawn from the analysis of Flesch scores. The scores for the various texts fell into four levels of readability, corresponding to the first four years of education (*very easy* - Flesch scores from 75 to 100), to the grades from the 5th to the 8th year (*easy* - scores from 50 to 75), to high school and college (*fairly difficult* - scores from 25 to 50) and to academic texts (*very difficult* - scores below 25). It is hoped that the tool presented here may open up the way for automatic screening of Portuguese texts according to the intended audiences.

Resumo

Este trabalho discute a aplicação de fórmulas de legibilidade a textos escritos em Português do Brasil. Trata-se de uma adaptação para o Português do conhecido índice de Flesch para o inglês, cuja fórmula foi modificada para refletir o fato de que as palavras do Português contém, em média, um número maior de sílabas do que as palavras do inglês. Testes similares foram feitos com outras fórmulas, tais como ARI, Kincaid e Coleman-Liau, sendo que os resultados levaram a conclusões análogas. Os índices para os vários tipos de textos classificam-se em quatro níveis de legibilidade, correspondendo aos primeiros quatro anos escolares (*muito fácil - 75 a 100*), da quinta à oitava série (*fácil - 50 a 75*), compreendendo o colegial e o nível superior (*pouco difícil - 25 a 50*), e textos acadêmicos (*muito difícil - abaixo de 25*). Testes exaustivos com textos reais dos mais variados tipos e níveis apontam para um desempenho muito satisfatório da fórmula adaptada, como pode ser observado pelos dados aqui apresentados.

I. Introduction

Readability formulas have been extensively used over the last fifty years for estimating the readability of written English texts (for a review see Klare, 1974-1975) with a number of different purposes. They are widely used for assessing the adequacy of remedial reading texts, military training manuals, elementary-school trade books, television captioning for the hearing impaired (Royer, 1990), and even for helping educational publishers design reading materials for target populations (Olson, 1986). There are more than 35 different formulas; in some of them word and sentence lengths are generally employed in an empirical equation (Olson, 1986) aimed at predicting how a particular kind of reader will perceive a text. Long words and sentences will obviously lead to more difficult texts. Some other formulas also consider word frequency and complexity, presence of complex noun phrases, Latin-based suffixes (Klare, 1974-1975). It must be pointed out that even in the most sophisticated formulas, what one tries to measure is the readability and not the comprehensibility of a given text (which can only be measured in tests with the readers themselves). There are a number of works showing a high correlation between readability scores and the degree of comprehensibility, but this has for a long time been a debatable issue (Klare, 1974-1975). Some authors go as far as discarding the effectiveness of the application of readability formulas altogether (Meade & Smith, 1991; Croll & Moskaluk, 1977; Mayo, 1993), and stress the potential harm such an application could bring. We shall delve into this discussion later on when analysing our own data.

The majority of studies of readability scores have been carried out for the English language. Recently, the application of some of these formulas has also become popular in software tools aimed at correcting style and grammar of English texts (CHERRY, 1982, MATZKIN, 1990). Because our research group is now involved in a project for developing this kind of tool for Brazilian Portuguese texts, we wondered whether readability formulas could be applied to this language. After an extensive search into the literature, we were unable to find any report on such studies. This has prompted us to investigate whether the formulas, originally developed for English, would also be meaningful for the analysis of Brazilian Portuguese texts. Several formulas, viz. Coleman-Liau, Kincaid, ARI and Flesch were employed in 120 passages of textbooks, and the subsequent analysis led to practically the same conclusions. We therefore concentrated on the results from the Flesch (1948) scores. Four levels of readability were identified, which correspond roughly to the different stages of Brazilian education.

II. Methodology and Results

The passages of text were randomly selected from a number of textbooks, covering the 8 years of primary education (grades 1-8), the high school (9-11), and introductory texts for college students (U). Ten passages of approximately 150 words each were used for each grade. The texts for the 1st to the 11th grades were extracted from Geography, History and Science books; whereas the college texts were obtained from introductory books of Humanities and Science. The academic texts were extracted from Humanities books. In order to have a more uniform set of passages, all of them were extracted from introductory parts of chapters or topics, and contained no mathematical equations, tables or figures. All textbooks are currently employed in official schools so that one can assume that their contents, especially the adequacy for the intended audience, may have gone through some degree of control.

The Flesch score was obtained from the equation [(Flesch, 1948; CHERRY, 1982)]:

$$\text{Reading Score} = 206.835 - 84.6*\text{syl.per.wd} - 1.015*\text{wds.per.sent}$$

where syl.per.wd is the average number of syllables per word and wds.per.sent is the average number of words per sentence.

Our first results showed that Flesch scores were systematically lower for Portuguese texts than for their English counterparts. A direct comparison was then made by calculating the Flesch score for original texts in English from Introductory Physics textbooks and also the Flesch score for the same texts translated into Portuguese. The translated textbooks employed have been used for a long time, and one can rely on the quality of the translation. A careful inspection indeed indicated that the translation was accurate. The Portuguese passages scored in average 42 points lower than the English passages. Therefore, for comparing the ranges of Flesch scores, we added 42 points to all Flesch scores obtained for the Portuguese texts analysed in this paper. It could be argued that our comparison was not extensive enough to warrant the generalization of this 42 point shift to all kinds of text. However, even if this shift differed from 42, it would not be to the extent of altering our conclusions in any substantial way. For example, in subsidiary experiments with a corpus of over 40 scientific reports we found out that English words have in average 1.8 syllables per word as compared to 2.2 syllables per word for the Portuguese words. This difference, caused mainly by the higher occurrence of Graeco-Latin words in

Portuguese, indeed leads to a shift of around 35 points in the Flesch scale.

Four different levels of readability scores could be identified as shown in Fig. 1 in which the Flesch scores (including the 42 point shift mentioned above) for the ten passages were plotted. For the sake of clarity, each passage for each academic year (from the 1st to College (U)) was numbered as indicated in the horizontal axis. These levels with Flesch scores in the ranges 100-75 (named **very easy**), 75-50 (**easy**), 50-25 (**fairly difficult**) and 25-0 (**very difficult**) correspond roughly to the stages of the Brazilian education which comprises three so-called degrees. Scores for the first degree which comprises the first 8 academic grades fall mainly into the two top ranges. In fact, in the Brazilian educational system this first degree is actually split into two phases, from the 1st to the 4th (children from 7 to 10 years old) and from the 5th to the 8th grades. The results from Fig. 1 seem to reflect that. The scores for textbooks of the 2nd degree that consists of 3 academic grades fell into the second lower range. The same happened with passages from introductory textbooks in the first years of college. In the lowest range for the **very difficult** texts only some points appeared. These arose from academic texts of areas in which there is a certain abundance of long technical words, with four or more syllables such as *aminoácidos* and *macromoléculas*. It must be pointed out, however, that these scores may have been deflated artificially by this high frequency of technical words. According to Cohen and Steinberg (1983), science texts are generally not as difficult to read as the formulas indicate because technical words are repeated so often that the readers soon get accustomed to them.

Other interesting features in Fig. 1 relate to the scores close to the borders of the different levels. There are few points in the border between the **very easy** and **easy** passages, probably because texts for the 5th to 8th grade students differ considerably from those for the first years of primary school. This difference may have arisen from the considerable break that occurs in the Brazilian educational methods in these stages. Up to the 4th grade a single teacher teaches all subjects to a group of students, whereas from the 5th grade on there is one teacher for each topic. The textbooks might obviously reflect these changes in method. The border between the **easy** and **fairly difficult** levels, on the other hand, is highly populated. Several texts fell into this border region, probably because there should not be significant differences in the students' reading ability at these stages.

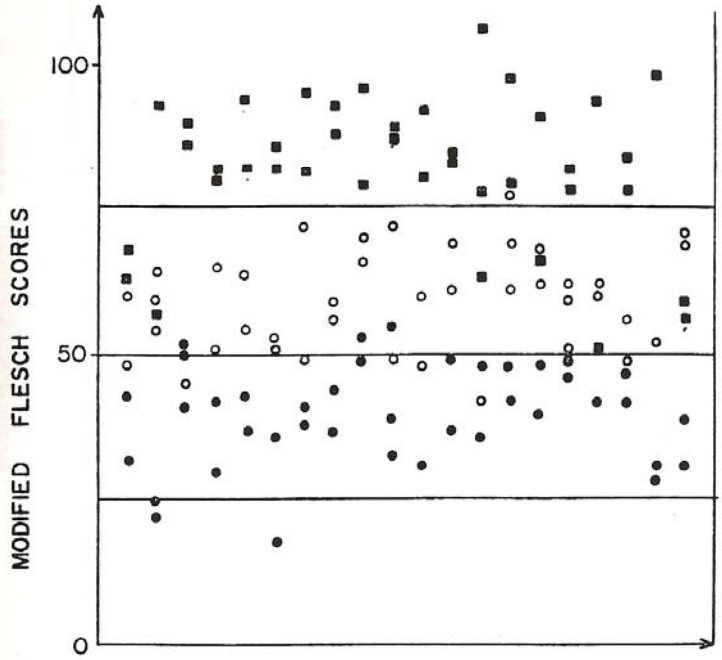

Fig. 1 - Modified Flesch scores for a number of texts. Points correspond, respectively, to: 1-4 grades: filled squares; 5-8 grades: open circles; 9-11 grades and college: filled circles

The evolution of the mean value of the Flesch score for each academic grade is shown in Fig. 2, where the error bars were calculated from the root mean square. One can observe that there is a significant decrease in the score for the first few grades when the alphabetization process is taking place. Not surprisingly, the changes in the subsequent grades (from the 5th to the 8th) are much smoother, until the scores eventually level off during the so-called 2nd degree (high school in the USA) and college. This leveling off was to be expected as some textbooks for the 2nd degree are classified according to the topic rather than to the academic grade they refer to. No substantial change should be observed in the readability of such books.

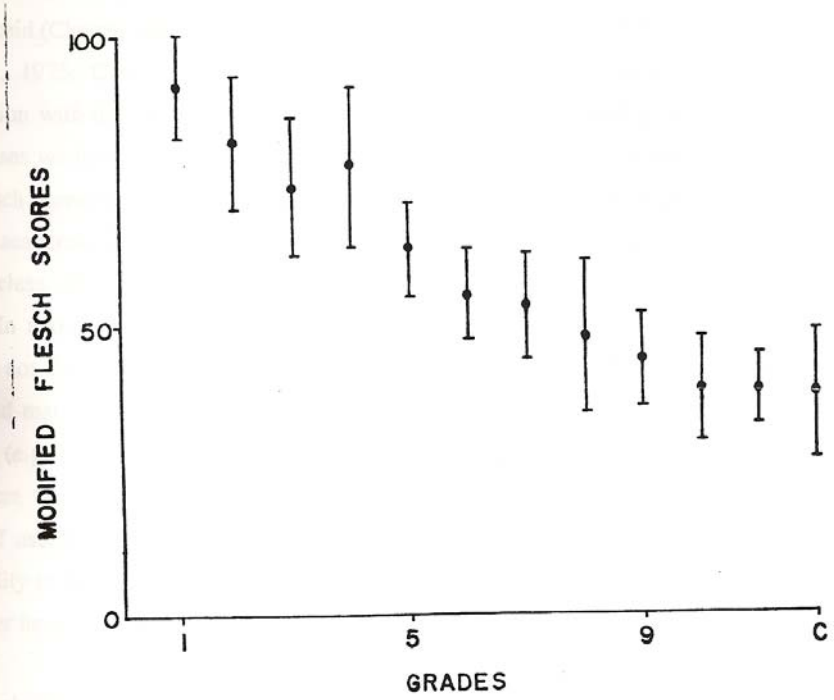

Fig. 2 - Modified Flesch scores plotted against the different grades in Brazilian education, from 1st to 11th grades and college (C). Each point corresponds to the average for each grade, and the error bars were calculated using the root mean square.

Approximately 20% of the passages tested did not fall into the category they were expected to, if the interval defined by the error bar is taken as a criterion. The size of the error bars varied from 12 to 22 points in the Flesch scale. Since the different levels of readability cover an interval of 25 points, we may safely say that the great majority of the passages are within the expected levels. There are few exceptions though, where passages had scores much lower than expected. A careful inspection in some of these *exceptional* passages showed that indeed they appeared to be rather difficult for the intended audience. Or, to say the least, they appeared to present more difficulties than the other passages for the same academic grade. This

was the case of some texts for the 1st to 4th grades.

In subsidiary tests, scores were calculated using other readability formulas. We employed the Kincaid (Cherry, 1982), ARI (Cherry, 1982; Klare, 1974-1975) and Coleman-Liau (Coleman & Liau, 1975; Cherry, 1982; Klare, 1974-1975) formulas and the results showed a high correlation with the Flesch scores, i.e. a text considered easy according to the Flesch scale in most cases would also be classified as easy by the other scores. We decided to concentrate on the Flesch scores because they follow a numerical scale from 0 to 100 rather than being related to the academic grades as in the other formulas. The results from the subsidiary tests, nevertheless, only confirm our general findings.

In another set of auxiliary experiments, we tested passages from quality newspapers (results not shown). Again distinct kinds of text fell into different levels of readability. Front page and main news fell into the **easy** category (Flesch scores from 75-50), while specialized articles (e.g. in scientific topics) generally fell into the **fairly difficult** category, especially if they were written by non-members of the newspaper staff. This happens because members of the staff usually follow very stringent rules in producing their text which tend to conform its readability to the newspapers standards. Passages from special sections devoted for children, on the other hand, scored in the **very easy** range as one should expect.

III. Analysis of the Results

The positive results presented in the last section appear to demonstrate the usefulness of readability formulas for Brazilian Portuguese texts. This success may have largely been achieved because the texts analysed were probably screened by educational authorities, in which content and other features were controlled in order to warrant adequacy for the target students. One could perhaps inquiry on whether the Flesch scores fell into the expected categories because the textbooks had already had their readability tested. This is very unlikely, for there has been no report of readability studies for Portuguese texts.

The results are quite conclusive, with different levels of readability being identified in real textbooks in accordance with what should be expected. However, it is obvious that no one can ensure that applying a formula based only on physical parameters of a text, like Flesch's formula, will warrant a desired degree of comprehensibility. In no way can we state, based on our results, that the texts which fell into a given category are actually adequate for the intended

audience. Unfortunately, there are no systematic studies on intelligibility of Portuguese texts which could confirm (or not) whether the estimated readability has a high correlation with the students' actual intelligibility. We may, nevertheless, elaborate upon Molina's work (Molina, 1979) on the intelligibility of Portuguese texts. In her investigation of intelligibility of literary texts by students from the 5th to 10th grades using the Cloze method, Molina observed that the degree of intelligibility was poorer when longer sentences were employed. This is consistent with the readability score being decreased with an increasing sentence length as in the Flesch formula.

Even though we are quite optimistic about the successful use of readability formulas for specific purposes, such as the one employed in this paper, there are however a number of issues to be considered. In the next section we highlight some of the limitations of the formula approach, also stressing the cases in which its application is most useful.

IV. Critique of the Formulas Approach

Before analysing the results presented in Section II we were for some time very skeptical about the efficacy of the formulas approach, in particular because of the strong criticisms put forward by a number of authors (Meade & Smith, 1991; Croll & Moskaluk, 1977; Mayo, 1993; Anderson & Davison, 1988; Rush, 1985; Olson, 1986). Amongst the many criticisms encountered in the literature, we have selected some points which we believe to be the most relevant ones. Flesch scores as well as related formulas do not reflect features that are relevant for comprehensibility, such as degree of abstraction, complexity and relationship of ideas, rethoric structure, and degree of discourse cohesion (Olson, 1986). Moreover, by using formulas one cannot discriminate between written discourse and nonsensical word combinations (Mayo, 1993). Major criticisms are also directed at the improper use of the formulas for assessing written materials. According to Rush (1985), they consistently predict that materials will be too difficult for the intended audience. Strong evidence of the misuse of readability formulas is provided especially in the case of adapted texts.

Another criticism which deserves special attention is the one put forward by Olson (1986). He pointed out that the McCall-Crabbs Standard Test Lessons in Reading - which were one of the sources employed in the derivation of the Flesch as well as other readability equations - did not vary in reading difficult to the extent assumed by those who developed the formulas. If this statement were absolutely correct, then no use of such formulas would be justifiable.

However, we carefully analysed Olson's results (Table 11 in his paper) and noticed that for the Flesch formula different levels of readability can actually be established, though there are intersections among the ranges defining these levels. In addition, in the derivation of the Flesch formula texts were also extracted from magazines for adult readers, which may have affected the factors in the formula (equation 1).

To counter all those criticisms one may argue that there is a substantial body of evidence that Flesch scores correlate well with the degree of comprehensibility of a given passage (as determined in actual tests with students). The success of the formulas approach has been proven in several specific cases [(Razik, 1969; Hill & Erwin, 1984; Grundner, 1978; Malouff et al., 1992)].

Because these formulas provide a quantitative way of estimating the difficulty of written material without testing readers, they may be used to quickly assess a wide range of texts (Olson, 1986), in particular with the help of a computer. This application, however, must be performed with extreme caution for it is irrefutable that some of the criticisms presented above are valid. In fact, this view is shared by authors such as Klare (1974-1975) who went as far as recommending when to use (or not to) the formulas. According to Klare (1984) [apud Rush, (1985): "Readability"] users of readability formulas should employ these formulas as a screening device, and consider that materials designed for training purposes are naturally more difficult than other kinds of text. He also stresses that formulas are poor predictors for high level texts where content is important, and urge users not to rely on formulas alone or use formulas as part of writing (Olson, 1986).

Alternative methods have been proposed in the literature for assessing the comprehensibility of given texts (Rush, 1985). One of such methods involves propositional analysis to determine the coherence of text passages (Clark, 1981 [apud Rush (1985): "Assessing Comprehensibility: The PHAN system"]). This method assesses readability by checking reference cohesion, use of connectives (e.g. because, then, after, etc.) and vocabulary difficulty. The obvious disadvantage of this method is the requirement of human expertise in the assessment of written materials, but - unlike readability formulas - it is appropriate for analysing texts to be employed with group of readers whose characteristics are known (Rush, 1985).

IV. Conclusions

In this article we showed that readability scores calculated with the Flesch formula are meaningful for Brazilian Portuguese texts. This is important insofar as applying a readability formula is extremely easy (with the help of a computer) and may serve as a quick way to screen textbooks and other written documents for their adequacy for a given audience. Also important is to bear in mind that formulas such as Flesch's are aimed at estimating readability and not comprehensibility. Other factors that are highly relevant for text comprehension are not contemplated in these readability formulas, and therefore readability scores should be used with extreme caution.

Acknowledgements

This project was funded by Itautec/Philco S.A. Financial assistance from FAPESP and CNPq is also acknowledged.

References

ANDERSON, R.C. & DAVISON, A. (1988). Conceptual and empirical bases of readability formulas. In: A. DAVISON & G.M. GREEN (Eds.), *Linguistic complexity and text comprehension: readability issues reconsidered*. (pp. 23-53), Hillsdale/NJ, Lawrence Erlbaum Associates.

CHERRY, L. (1982). Writing tools. *IEEE Transactions on communications, COM-30*, 100-105.

COHEN, S.A. & STEINBERG, J.E. (1983). Effects of three types of vocabulary on readability of intermediate grade science textbooks: An application of Finn's transfer feature theory. *Reading Research Quarterly, 19*, 87-101.

COLEMAN, M. & LIAU, T.L. (1975). A computer readability formula for machine scoring. *Journal of Applied Psychology, 60*, 283-284.

CROLL, W. & MOSKALUK, S. (1977). Should Flesch counts count? *Teaching of Psychology, 4*, 48-49.

FLESCH, R.F. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*, 221-233.

GRUNDNER, T.M. (1978). Two formulas for determining the readability of subject consent

forms. *American Psychologist,* 773-775.

HILL, W.R. & ERWIN, R.W. (1984). The Readability of content textbooks used in middle and junior high schools. *Reading Psychology: An International Quarterly, 5*, 105-117.

KLARE, Feorge R. Assessing readability. *Reading Research Quarterly*. N. 1, 1974-1975, X/1 (p. 62-102)

LANDRIGAN, David T. & PALLADINO, Joseph J. A reply to Gillen. *American Psychologist*. July, 1974 (p. 571-2)

MALOUFF, John et al. Readability of health warnings on alcohol and tobacco products. *American Journal of Public Health*, V. 82, N. 3, March, 1992 (p. 464)

MATZKIN, J. Grammar Checkers improve, but won't replace your English teacher. *PC Magazine*. March, 13:46 (1990)

MAYO, Donna J. Beware the jabberwock and other perils of readability formulae. *Transfusion*. V. 33, N. 10, 1993 (p. 884-5)

MEADE, Cathy D. & SMITH, Cyrus. Readability formulas: cautions and criteria. *Patient Education and Counseling*, V. 17, 1991 (p. 153-8)

MOLINA, Olga. *A avaliação da inteligibilidade de livros didáticos de 1º e 2º graus por meio da técnica Cloze*. PhD. Thesis, Instituto de Psicologia, São Paulo, 1979 - in Portuguese

OLSON, Arthur V. A question of reading validity. *Journal of Research and Development in Education*. V. 19, N. 4, 1986 (p. 33-40)

RAZIK, Taher A. A study of American newspaper readability. *The Journal of Communication*, V. 19, December, 1969 (p. 317-324)

ROYER, James M. The readability dilemma. *Contemporary Psychology*, N. 1, V. 35, 1990 (p. 41-2)

RUSH, R. Timothy. Assessing readability formulas and alternatives. *The Reading Teacher*. Dec/1985 (p. 174-283)