



Revista Universo Contábil, ISSN 1809-3337  
Blumenau, v. 16, n. 2, p. 110-126, abr./jun., 2020

doi:10.4270/ruc.2020209  
Disponível em [www.furb.br/universocontabil](http://www.furb.br/universocontabil)



## INTELLIGIBILITY VS READABILITY: UNDERSTANDABILITY MEASURES OF FINANCIAL INFORMATION

### INTELIGIBILIDADE VERSUS LEGIBILIDADE: MEDIDAS DA COMPREENSIBILIDADE DAS INFORMAÇÕES FINANCEIRAS

### INTELIGIBILIDAD VS LEGIBILIDAD: MEDIDAS DE COMPRENSIÓN DE LA INFORMACIÓN FINANCIERA

Recebido em: 11-02-2020  
Avaliado em: 20-04-2020  
Reformulado em: 06-05-2020  
Aceito para publicação em: 28-05-2020  
Publicado em: 26-04-2021  
Editor Responsável: Roberto Carlos Klann

**Samantha Valentim Telles<sup>1</sup>**

**Bruno Meirelles Salotti<sup>2</sup>**

#### ABSTRACT

Decision-making is influenced by information, which must be understood so it can be useful. However, measuring understandability is not a simple task. Previous studies use readability metrics as proxies for understandability, but readability is a shallow metric ignoring discourse and language components. This research analyzes the notes to the financial statements using intelligibility metrics, an alternative to readability measures. Our sample comprised 44 Brazilian firms that presented their Portuguese notes (native language) and English (non-native) from 2012 to 2015. Focusing on the notes of Financial Instruments and Provisions, we found that, for most indexes where firms had worse readability levels, they showed better intelligibility levels. It indicates that both metrics measure different things. Our results also indicated that language impacts these metrics and that firms did not improve their information quality after the guideline OCPC 07 from the Brazilian Accounting Standards Committee.

**Keywords:** Readability. Intelligibility. Financial Information. Notes to the financial statements.

<sup>1</sup>Doutora em Controladoria e Contabilidade pela Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo; Professora da Faculdade FIPECAFI; E-mail: [samanthatelles@outlook.com](mailto:samanthatelles@outlook.com)

<sup>2</sup>Doutor em Controladoria e Contabilidade pela Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo; Professor da FEA/USP; E-mail: [brunofea@usp.br](mailto:brunofea@usp.br)

## RESUMO

A tomada de decisão é influenciada pelas informações, que devem ser compreensíveis para que possam ser úteis. No entanto, medir a compreensibilidade não é uma tarefa simples. Estudos anteriores usam métricas de legibilidade como *proxies* para a compreensibilidade, mas a legibilidade é mostrada como uma métrica superficial, que ignora os componentes do discurso e da linguagem. Esta pesquisa analisa as notas explicativas usando métricas de inteligibilidade, uma alternativa às medidas de legibilidade. A amostra foi composta por 44 empresas brasileiras que apresentaram suas notas tanto em português (idioma nativo) quanto em inglês (idioma não nativo) de 2012 a 2015. Ao focar nas notas de Instrumentos e Provisões Financeiras, descobriu-se que, para a maioria dos índices em que as empresas tinham piores níveis de legibilidade, elas mostraram melhores níveis de inteligibilidade. Isso indica que ambas as métricas medem coisas diferentes. Nossos resultados também indicaram que o idioma afeta essas métricas e que as empresas não melhoraram a qualidade de suas informações após a OCPC 07 do Comitê de Pronunciamentos Contábeis.

**Palavras-chave:** Legibilidade. Inteligibilidade. Informações Financeiras. Notas Explicativas.

## RESUMEN

La toma de decisiones está influenciada por la información, que debe ser comprensible para que pueda ser útil. Sin embargo, medir la comprensibilidad no es una tarea simple. Estudios anteriores utilizan métricas de legibilidad como indicadores de la comprensibilidad, pero se muestra que la legibilidad es una métrica superficial, que ignora los componentes del discurso y del lenguaje. Esta investigación analiza las notas explicativas utilizando métricas de inteligibilidad, una alternativa a las medidas de legibilidad. Nuestra muestra comprendió 44 empresas brasileñas que presentaron sus notas tanto en portugués (idioma nativo) como en inglés (idioma no nativo) de 2012 a 2015. Centrándonos en las notas de instrumentos financieros y provisiones, encontramos que, para la mayoría de los índices donde las empresas tenían peores niveles de legibilidad, mostraron mejores niveles de inteligibilidad. Esto indica que ambas métricas miden cosas diferentes. Nuestros resultados también indicaron que el lenguaje impacta estas métricas y que las empresas no mejoraron la calidad de su información después de la directriz OCPC 07 del Comité de Pronunciamentos Contables.

**Palabras-clave:** Gobernanza corporativa. Instituciones de Enseñanza Superior. Gobernanza en Instituciones de enseñanza superior.

## 1 INTRODUCTION

The Conceptual Framework of the Brazilian Accounting Standards Committee (Comitê de Pronunciamentos Contábeis – CPC) says that useful financial information must be, among other things, understandable. Previous studies showed that firms with poor performances might write longer reports to avoid litigation or make them more costly to analyze (Bloomfield, 2008; Li, 2008).

At the end of 2014, the CPC published a new guideline, the Brazilian Accounting Standards Guideline number 07 (OCPC 07). This guideline presents the basic requirements for preparing and disclosing financial reports, especially to the notes. One of the requirements is understandability; thus, firms are expected to improve this characteristic of their statements and notes. Moreover, a study of the auditing firm EY, which monitors the comments on public company filings from the Securities and Exchange Commission (SEC), found that many comments demand better disclosure to improve the understandability of financial statements' information (EY, 2016).

Nevertheless, there is not a direct way for companies to measure such understandability. Overall, readability and understandability have been treated as synonymous. However, besides being different concepts, readability metrics are very simplistic: they rely on shallow metrics such as sentence and word lengths and ignore many other discourse and language components that, as theory

indicates, have an impact on comprehension difficulty (Lewis, Parker, Pound, & Sutcliffe, 1986; Jones, 1997; Graesser, McNamara, Louwerse, & Cai, 2004).

Readability formulas measure factors that could be correlated with difficulty, but they cannot point to all text features that may impact comprehension. They neglect ideas complexity, sentence and paragraph construction, individual differences, and discourse markers. Such formulas cannot assess more deeply the reasons and correlations of factors that make the text difficult to understand (Razik, 1969; Dreyer, 1984; Scarton, Almeida, & Aluísio, 2010). As an example of such disconnection, Schriver (1989) mentions that a text written backward would have the same readability score as the one in regular order. However, the understanding would be much more difficult.

To predict comprehension a metric would have to include a reader's knowledge, language abilities, and other cognitive characteristics, which are not considered by readability formulas (McNamara, Louwerse, & Graesser, 2002). Consequently, we need to search for a new way of measuring understandability.

The Coh-Metrix software has tools that measure the overall cohesion and difficulty of a text; it can provide scores of several cohesion features. It determines the coherence or adequacy of a text to a reader (McNamara et al., 2002, Scarton et al., 2010). Therefore, this research aims to analyze the intelligibility and readability relation on notes to the financial statements.

In Brazil, a group of researchers from the University of Sao Paulo developed a version of the original Coh-Metrix adapted to Portuguese, the Coh-Metrix-Port (Finatto, 2011). Thereby, the specific purposes of our research are: (i) to compare the results of intelligibility and readability metrics; (ii) to verify if the levels of readability and intelligibility changed after OCPC 07; and (iii) to verify the impact of the reporting language.

Most previous studies regarding readability focus solely on reports in English (Moreno & Casasola, 2015). Besides including an analysis of information in Portuguese, we introduce a new way of measuring understandability through intelligibility metrics, which researchers in linguistics have used. The comparison between both metrics and between these two languages has not been made yet for Financial Information.

## **2 THEORETICAL FRAMEWORK**

### **2.1 Readability**

Readability and understandability have different concepts. Readability corresponds to the complexity of the display. It is passive and text-centered, measuring a passage's textual difficulty. Understandability is the users' ability in discerning proper meaning of a passage, considers text-reader interaction. It also depends on a syntactical difficulty and the readers' characteristics, such as context, background, prior knowledge, education, interest, and general reading ability (Smith & Taffler, 1992; Jones, 1997; Jones & Smith, 2014).

We found many papers about the readability of financial information in EBSCOHost from the end of 1960's until the beginning of 1980's. In the late 2000's the theme began to grow stronger again in many international journals, but with different focuses. The first time it became an important subject, many papers studied the relation between accounting information and the Theory of Information. Recently, studies focus on what influences such types of disclosures. Recently, it is not a common subject in papers abroad, but some researches (Gomes, Ferreira, & Martins, 2018; Santos, Calixto, & Bispo, 2019) may be indicating a new wave of readability studies now in Brazil.

Researchers developed formulas with the intent of creating indexes of probable difficulty. Nonetheless, they are strictly text-based, and they do not explore factors regarding meaning communication (Dreyer, 1984; Rush, 1985). Most metrics use two components, one that measures sentence length and another which measures word complexity (Table 1).

**Table 1 – Readability indexes applied to financial reports**

<b>Metrics</b>	<b>How it works</b>	<b>Researches in Accounting</b>
Flesch	Calculated by: - Number of syllables per word - The average size of sentences	Soper and Dolphin Junior (1964) Smith and Smith (1971) Lewis et al. (1986) Smith and Taffler (1992) Kumar (2014) Moreno and Casasola (2015) Gomes et al. (2018)
Dale-Chall	Calculated by: - Number of complex words (words that are not present in a list of words made by Dale) - The average size of sentences	Smith and Smith (1971) Lewis et al. (1986)
Gunning Fog Index	Calculated by: - Number of complex words (words with 3 or more syllables) - The average size of sentences	Lewis et al. (1986) Li (2008) Lehavy, Li, and Merkley (2011) Lawrence (2013) Lundholm, Rogo, and Zhang (2014) Merkley (2014) Loughran and McDonald (2014) Santos et al. (2019)
Kwolek	Calculated by: - Number of complex words (words with more than three syllables, abbreviations, or symbols) - The average size of sentences	Lewis et al. (1986)
Lix	Calculated by: - Number of long words (words with more than six letters) - The average size of sentences	Lewis et al. (1986) Smith and Taffler (1992)

Soper and Dolphin Junior (1964) found through an experiment with four different groups of people that a relation exists between understandability and readability. Similarly, Smith and Taffler's (1992) research sought to provide evidence of the differences between readability and understandability. To do so, they used two metrics to measure readability, Flesch, and Lix, while cloze measured understandability. They found that for more sophisticated users, such as accountants, Flesch and cloze scores were much closer than to a non-sophisticated group like undergraduate students.

Previous studies (Smith & Smith, 1971; Worthington, 1977; Lewis et al., 1986; Kumar, 2014) indicated that financial information is classified as difficult. Kumar (2014), for instance, found that none of the studied companies were providing annual reports classified as easy to read. Moreno and Casasola (2015) analyzed the readability evolution of annual reports' narratives in Spanish using an adapted version of the Flesch index. They verified that readability improved with time. Both mentioned researches were one of the few readability studies in non-English accounting documents.

Lundholm et al. (2014) studied lexical and numerical properties of foreign companies' communication listed in the United States from 2000 to 2012. They verified that foreign companies with stocks in the U.S. write more readable texts in their Management Discussion and Analysis (MD&A) and press releases than American firms, to attract more American investment.

Many other studies looked at readability, not analyzing its relation with readability, but mostly to what determines it or what it affects. Examples of that were Li (2008), Lehavy et al. (2011), Lawrence (2013), Loughran and McDonald (2014), Merkley (2014), Guay, Samuels, and Taylor (2016) and Rennekamp (2012), this last one faced the topic using an experiment.

## 2.2 Intelligibility

Usually, passages that are difficult to read are also challenging to understand, but it is not possible to say that both concepts are perfect correlates (Jones, 1997). The act of reading can be divided into three factors to allow understanding: (a) the role of the reader, (b) the role of the text, and (c) the process of interaction between reader and text (Leffa, 1996). Among the factor related to the text, Scarton et al. (2010) highlight readability (the graphical representation of the text) and intelligibility (the use of frequent words and less complex syntactic structures).

To Scarton et al. (2010), some factors help improving understandability, like cohesion and coherence. These two concepts represent how words and ideas informed relate to certain levels of discourse, language, and global knowledge (McNamara et al., 2002). While cohesion is a text characteristic, coherence is a characteristic of a reader's mental representation of such text (Graesser et al., 2004).

In cohesion, the connections are based on explicit linguistic elements, such as words, characteristics, signals, and others, and how they are combined. These elements are interpreted in a sociocultural context and guide the reader in interpreting the text's substantive ideas, making connections between different ideas, and making connections of these ideas with its topics or themes (McNamara et al., 2002; Graesser et al., 2004).

Coherence is the text characteristics (i.e., cohesion aspects) that must contribute to the coherence of the reader's mental representation. It is a result of the interaction between reader and text. The mental representation connections are based on elements available in the text and the reader's abilities, knowledge, and intentions. The reader will form this coherent mental representation of the text if he/she has enough world knowledge regarding the subject and if the linguistic and discourse cues are also enough (McNamara et al., 2002; Graesser et al., 2004; Scarton et al., 2010).

It is possible to conclude that a text will be understandable if it is coherent and has cohesion. However, there is nothing the writer can do to improve cohesion if the reader has no world knowledge of such subject, once coherence depends on the characteristics of the reader and not only on the text. Considering the above, this research assumes that intelligibility is the understandability that does not depend on the user.

Researchers of the Institute for Intelligent Systems at the University of Memphis developed a system called Coh-Matrix, or coherence/cohesion metrics, to analyze cohesion. This system is a tool used to compute computational coherence and cohesion metrics (Coh-Matrix Website, 2017). One can use this tool to measure the overall cohesion and text's difficulty (vocabulary issues, syntactic composition, and meaning), improving current readability indexes. It can provide scores of several cohesion features (lexical, syntactic, discourse, and conceptual). It determines the coherence or adequacy of a text to a reader (McNamara et al., 2002; Scarton et al., 2010).

Among the measures of Coh-Matrix are word frequency, density scores, logical operators, count of connectives, readability, among others. Over 500 metrics are available in a restricted version of Coh-Matrix; only 60 are available at the online version, and 48 of those 60 metrics were adapted to Portuguese at the Coh-Matrix-Port (Finatto, 2011).

## 2.3 Hypotheses

In accordance to Worthington (1977), it is possible to establish a writing style that is appropriate to the average investor. Overall, readability indexes try to create proxies for possible problems of misinterpretation and mislead decision-making. Dale-Chall model uses a set of words in which words outside it are considered complexes. Lix index uses the number of letters as a proxy to complexity, while in Flesch, FOG, and Kwolek techniques, the proxy is the number of syllables, the greatest the number of syllables, the more complex it is.

However, the readability metrics are limited. The use of readability formulas by textbook writers, for instance, makes them reduce sentences to improve their readability scores, but it usually

has a negative impact on texts (Dreyer, 1984; Graesser et al., 2004). By doing so, the texts may end up with cut and short sentences with low levels of cohesion, maybe becoming even more difficult to understand, especially in cases where the reader does not have much knowledge on the subject or low reading proficiency (Dreyer, 1984; McNamara et al., 2002). However, if this change is done maintaining the coherence of a text, it is probable that, understandability is improved. Thus, the level of readability affects, significantly, the degree of comprehension, but there might not be a cause-effect relation here (Razik, 1969; Dreyer, 1984; Graesser et al., 2004).

Once intelligibility metrics measure the cohesion of a text, considering the coherence constant, our first hypothesis is:

*H1: High readability texts also have high intelligibility.*

Most literature related to readability and understandability, especially in Accounting, is based on texts in English. Klare (1974-1975) showed that, once most readability formulas were developed to passages written in English, the application to other languages, in most cases, demands an adaptation in readability formulas. Martins, Ghiraldelo, Nunes, and Oliveira Junior (1996) developed an adaptation of the Flesch index to Portuguese. However, we understand that even if the formula is adapted, the scores cannot be the same for different languages.

The study of Pasqualini, Scarton, and Finatto (2011) shows differences in the intelligibility of texts in Portuguese and English, which, they believe, may be due to their grammar, which functions differently in each language. Translation problems are even worse for Accounting because it relies on specialized, culture-specific terminology (Evans, Baskerville, & Nara, 2015). On the other hand, Lundholm et al. (2014) verified that firms usually compensate for geographic distance with clearer disclosures. Either way, it is possible to expect that reporting language (native versus non-native language) impacts intelligibility. Thus, we created the following hypotheses:

*H2A: Reporting language has an impact on notes' readability.*

*H2B: Reporting language has an impact on notes' intelligibility.*

Previous studies, such as Barth, Landsman, and Lang (2008), showed that applying the International Financial Reporting Standards (IFRS) improves information quality. Brazil already applies the IFRS since 2010, but one should expect that firms will continue improving their information, especially when a guideline such as the OCPC 07 is published. To improve information, the OCPC 07 required, among other things, a better understandability of financial statements, including the notes. Therefore, one should expect that after it was published, firms would improve such characteristics of their information. Santos et al. (2019) studied the impact of OCPC over size, readability, and notes specificity. They analyzed notes from 2010, 2013, 2014, and 2015. They found a decrease of 10% of the notes' size among firms from B3's Novo Mercado or those audited by a Big Four. However, they could not find any significant increase in readability, measured using the Fog index. Similarly, Gomes et al. (2018), using the Flesch index to measure readability, could not find any improvement after OCPC 07.

Even though previous evidence does not show any impact on readability, they studied different years or used different methods. If we consider that readability and intelligibility are good proxies for understandability, we should expect it to improve over time (as at Moreno and Casasola, 2015), especially after the guideline. Then, we created the last hypothesis:

*H3: The intelligibility and readability of notes to the financial statements improved after OCPC 07.*

### 3 METHODOLOGY

In general, accounting academics employed readability formulas and comprehension tests to assess accounting narratives' effectiveness regarding both readability and understandability (Jones, 1997). According to Smith and Smith (1971), notes' readability analysis is an objective metric to measure a message's attribution level of accounting reports' intended meaning. As previously discussed, the intelligibility measures are better proxies once they consider the cohesion of the text.

We studied Brazilian public firms listed in American and Brazilian stock exchanges between 2012 and 2015. To test H1, we separated the firms of our sample into quartiles. Then, through mean tests and equivalents, we verified how different readability levels are related to each of the multiple intelligibility indexes analyzed.

To test H2, we verified if a company is more readable/intelligible or not when presenting in their local language. The comparison with the United States is something that was not done before. To test the level of similarity in between metrics, we used mean tests or equivalents for nonparametric cases to verify if they have the same scores, on average.

To test H3, we analyzed if the scores changed from 2012 to 2015, as expected, given the publication of OCPC 07 at the end of 2014.

The time section chosen for this study was due to the OCPC 07. We did not consider that a larger period would have changed our results significantly.

We could only analyze those firms that had information for all four years both in English and Portuguese, and those that the conversion of documents from pdf to Word® were possible. The remaining sample comprised 44 firms, i.e., a total of 176 firm-year observations. Table 2 presents the sample.

**Table 2 – Research sample**

<b>Initial sample (Brazilian firms that have presented information in Portuguese and English)</b>	<b>72</b>
We could not find the financial statements for at least one of the languages and in at least one of the years	(18)
We could not convert at least one statement from pdf to Word®	(10)
<b>Final sample</b>	<b>44</b>

The exclusions enabled us to perform a paired average test. Thus, we controlled for firms characteristics by using the same companies for both languages.

To calculate the cohesion and Flesch indexes, we used the software Coh-Metrix to analyze notes in English, and the Coh-Metrix-Port, to analyze notes in Portuguese. Coh-Metrix-Port calculates the readability of each firm's notes for Brazilian Portuguese, using Martins et al.'s (1996) formula.

Coh-Metrix and Coh-Metrix-Port are not very specific about how they separate sentences. We tested it and verified that they do not consider colons and semicolons as sentence separators. So, we removed all commas and converted all colons and semicolons in periods. Similarly, instead of removing headings, as Loughran and McDonald (2014), we put periods in it. This way, the end of the sentence would always be the periods, making all reports comparable. Using those premises allows firms that present their information in topics to have better readability scores, for instance.

The numbers, including dates, were also excluded from the analysis, as in Loughran and McDonald (2014). Abbreviations such as 'a.a.', 'p.a.', 'e.g.' and 'i.e.' were also eliminated. We excluded all dashes and parentheses once the comments they present may be considered part of the sentence. We removed tables, so the only written text would be analyzed.

Given that Coh-Metrix cannot process texts larger than 15,000 characters, we compared some extractions of specific notes (Operations/Corporate or General Information, Accounting policies/practices, Financial Instruments/Risk Management, Provisions/Contingencies, and Equity) with the whole text in Portuguese, using mean tests. We found that the notes which readability and intelligibility scores resemble the most with the full text were Financial Instruments/Risk

Management and Provisions/Contingencies, which we called Financial Instruments and Provisions, in this paper, respectively.

Financial Instruments is not different from the full report, statistically, in 19 of the 49 indexes, at a 5% significance level, while Provisions is not statistically different from the full notes in 24 indexes. Together, both notes are equal to the full report in 34 different indexes.

Only in a few indexes Operations, Policies, and Equity (Ambiguity of adjectives, Ambiguity of verbs, Temporal connectives, and Temporal negative connectives) were equal to the full report when Financial Instruments and Provisions were not. Thereby, including any of these three notes in this study does not show much value because the chosen notes already capture most of the similarity to the full note.

### 3.1 Analyzed Variable of Readability

Most readability metrics work to evaluate sentences in English. We could not find any adaptation to Portuguese of other most commonly used techniques such as Dale-Chall and Gunning Fog. Thus, we chose to use the Flesch index because it was adapted to Portuguese by Martins et al. (1996). The Flesch index considers two factors: the number of syllables and the average size of a sentence. Thus, the bigger the sentences and/or the number of syllables per word, the worse is the score given to accounting reports, and the higher the score, the better the readability (Soper & Dolphin Junior, 1964; Smith & Smith, 1971).

Such index was created based on the textbooks written complexity of some disciplines of different levels of education (middle school, high school, and college), according to the number of syllables in words and number of words per sentence. The Portuguese adaptation considered that words in Portuguese usually have more syllables, leading to a shift of approximately 42 points in the Flesch scale. Therefore, in the formula adapted to Portuguese by Martins et al. (1996), instead of using a constant of 206.835, it uses 248.835. The formulas we used to evaluate the readability of passages were those of Table 3.

**Table 3 – Flesch formulas: English vs. Portuguese**

Original Flesch Index	Adapted Flesch Index (Brazilian Portuguese)
– 84.6 (Number of syllables/Number of words)	– 84.6 (Number of syllables/Number of words)
– 1.015 (Number of words/ Number of sentences)	– 1.015 (Number of words/ Number of sentences)
+ 206.835	+ 248.835

Table 4 presents the classification of readability levels for each Flesch index.

**Table 4 – Flesch scale comparison: English vs. Portuguese**

English (Original)		Portuguese (Martins et al, 1996)	
Score	Difficulty level	Score	Difficulty level
<b>0 – 30</b>	Very Difficult	<b>70 – 80</b>	Fairly Easy
<b>30 – 50</b>	Difficult	<b>80 – 90</b>	Easy
<b>50 – 60</b>	Fairly Difficult	<b>90 – 100</b>	Very Easy
<b>60 – 70</b>	Standard	<b>75 - 100</b>	Very Easy

### 3.2 Studied Metrics

There are some differences in the calculations of Coh-Metrix and Coh-Metrix-Port. As Pasqualini et al. (2011), one cannot compare all metrics because each software has its metrics, and some are not compatible given the source of the data used to calculate the indexes. Thus, we only considered the variables calculated in the same way in both languages. Table 5 present the variables



that represent 28 readability and intelligibility indexes; out of those, 15 indexes of Financial Instruments and Provisions were significantly equal to the full note.

**Table 5 – Variables chose for this study**

Readability indexes		Intelligibility indexes
Flesch index	Adjective incidence	Connectives incidence <sup>1 2</sup>
Mean sentences / paragraph <sup>1</sup>	Noun incidence <sup>2</sup>	The incidence of additive connectives <sup>1 2</sup>
Mean syllables/word	Adverb incidence <sup>1</sup>	The incidence of causal connectives <sup>2</sup>
Mean words / sentence <sup>1</sup>	Pronoun incidence	The incidence of logical connectives <sup>1 2</sup>
Number of Paragraphs	Verb incidence	The incidence of temporal connectives
Number of Sentences	Mean hypernyms/verb	Adjacent argument overlap <sup>1</sup>
Number of Words	Personal pronouns incidence <sup>2</sup>	Argument overlap
	Type to token ratio	Adjacent stem overlap <sup>1</sup>
	Modifiers / Noun Phrase <sup>2</sup>	Stem overlap
	Noun Phrase Incidence <sup>2</sup>	Adjacent content word overlap <sup>1</sup>
	Words before Main Verb <sup>1 2</sup>	

Note. <sup>1</sup> Variables that the note of Financial Instruments was indistinguishable from the full note, in a 5% significance level; <sup>2</sup> Variables that the note of Provisions was indistinguishable from the full note, in a 5% significance level.

Flesch components are word length, represented by “Mean syllables per word”, and sentence length, calculated as “Mean words per sentence”. Other indexes presented in Table 5 related to readability were: “Mean sentences per paragraph”, “Number of Paragraphs”, “Number of Sentences”, and “Number of Words”. All others can be considered indexes for intelligibility.

## 4 RESULTS

### 4.1 Readability vs. Intelligibility

According to the classification of Flesch of Martins et al. (1996), we found that for notes in English, 74.4% of our sample was classified as “Very difficult” for the note of Financial Instruments, while the rest is considered “Difficult”. This proportion is not much different from the sample of Smith and Smith (1971), where 73.5% was “Very difficult”, and 26.5% was “Difficult”. On the other hand, for Provisions' note, 54.5% was “Very difficult”, and 45.5% was “Difficult”.

For notes in Portuguese, the results are similar for Financial Instruments but much different for Provisions. The percentage of “Very difficult” was 72.7% for the first and 93.2% for the second. The rest was classified as “Reasonably difficult”, as the classification of Portuguese diverges from the one in English.

To analyze how different readability levels impact intelligibility, we separated the sample into quartiles according to its Flesch score. Then we tested if the intelligibility indexes of the lowest (1) and highest (4) quartiles could be considered indistinguishable. The first quartile is the one with the worse levels of readability, while the other comprises the best levels. We tested the Flesch index in both to make sure that they were statistically different, and in all cases, we confirmed: all Sigmas were of < 0.001 in both parametric and nonparametric tests.

We performed the means tests according to the normality assumptions of the parametric t-test (normal distribution and homogeneous variance). In cases where it was rejected, we used the Mann-Whitney test for independent samples. We presented the results of such tests in Table 6.

It is possible to verify that the results are slightly different for each scenario, but it is almost unanimous that, in a 5% significance level, the Overlaps are all different from one quartile to the other. Overall, the quartile with the worse readability showed the highest means in all of these variables. It means, however, that this quartile has more repetition of words, stems, and arguments, which should make the text easier to understand.

**Table 6 – Readability vs. Intelligibility: Quartiles**

Variables	Financial Instruments				Provisions			
	Portuguese		English		Portuguese		English	
	Test	Sigma	Test	Sigma	Test	Sigma	Test	Sigma
<b>Connectives</b>								
All connectives	t-test	0.565	t-test	0.023	Mann	0.002	Mann	0.565
Additive connectives	t-test	0.208	t-test	0.005	Mann	0.071	Mann	0.003
Causal connectives	t-test	0.116	Mann	0.793	Mann	0.003	Mann	0.367
Logical connectives	t-test	0.072	Mann	0.070	Mann	0.004	t-test	0.016
Temporal connectives	t-test	0.030	t-test	0.458	t-test	0.373	Mann	0.105
<b>Incidences of Content Words and Personal Pronouns</b>								
Adjective	Mann	0.000	Mann	0.010	Mann	0.000	Mann	0.537
Adverb	t-test	0.207	t-test	0.043	t-test	0.000	Mann	0.385
Noun	t-test	0.029	Mann	0.967	t-test	0.001	Mann	0.802
Pronoun	Mann	0.980	t-test	0.633	Mann	0.276	t-test	0.851
Verb	Mann	0.676	Mann	0.443	Mann	0.243	t-test	0.013
Personal pronouns	Mann	0.388	Mann	0.179	Mann	0.017	Mann	0.520
<b>Overlaps</b>								
Adjacent argument	Mann	0.007	Mann	0.003	Mann	0.002	t-test	0.008
Argument	Mann	0.101	Mann	0.000	Mann	0.007	t-test	0.006
Adjacent stem	Mann	0.000	t-test	0.001	Mann	0.000	t-test	0.002
Stem	Mann	0.000	Mann	0.000	Mann	0.001	t-test	0.002
Adjacent content word	Mann	0.003	t-test	0.894	Mann	0.000	Mann	0.066
<b>Other Indexes</b>								
Mean hypernyms per verb	Mann	0.176	t-test	0.060	t-test	0.468	t-test	0.152
Type to token ratio	Mann	0.049	t-test	0.692	Mann	0.000	Mann	0.496
Modifiers per Noun Phrase	Mann	0.068	t-test	0.484	Mann	0.005	t-test	0.017
Noun Phrase Incidence	Mann	0.739	t-test	0.111	t-test	0.017	Mann	0.418
Words before Main Verb	t-test	0.005	t-test	0.055	t-test	0.043	Mann	0.234

Note. Mann = Mann-Whitney test. The Sigmas correspond to a two-tailed test.

Connectives organize a text, which should facilitate reading. In all cases where the means were statistically different, the first quartile showed the highest incidence of connectives, which also went against the expected.

On “Incidences of Content Words and Personal Pronouns”, in five out of the seven cases with differences between quartiles, the first one (Adverb) showed higher means. Nonetheless, the understandability will depend on the text as a whole.

On “Mean hypernyms per verb” only for Financial Instruments (English), the quartiles showed some difference. However, the first one had the higher mean, and higher levels of this index mean that words are considered more concrete (Graesser et al., 2004), i.e., easier to be understood.

“Noun Phrase Incidence” for Provisions in Portuguese is also statistically different. However, the higher mean belongs to quartile 4. It suggests that this quartile is informationally dense and has more complex syntax (Coh-Metrix-Port Website, 2017) in comparison to the other.

The other variables with a difference statistically significant between quartiles showed the expected behavior, i.e., the greater the readability, the greater the intelligibility.

Overall, these results show that when the two quartiles do not have the same understandability, the easier to understand is the first, which has lower readability levels, indicating a disconnection between these two measures.

A limitation to this analysis was that many times, a single firm was considered multiple times in the same quartile. However, the composition of each quartile changed from one language/note analyzed to the other. Therefore, there should be no significant impact on our results regarding this limitation.

## 4.2 Portuguese and English Comparison

We also tested if both languages had the same scores. In Table 7, we tested if both languages' scores were, on average, statistically equal (using t-test or Wilcoxon, according to normality tests of differences).

**Table 7 – Mean/Median tests Portuguese vs. English**

Variables	Financial Instruments		Provisions	
	t/z	95% CI	t/z	95% CI
<b>Panel A: Readability Indexes</b>				
Flesch index	z = -10.801***	Different	z = -11.505***	Different
Mean sentences per paragraph	z = -3.141***	Different	z = -2.857***	Different
Mean syllables per word	z = -11.505***	Different	z = -11.505***	Different
Mean words per sentence	z = -11.170***	Different	z = -7.095***	Different
Number of Paragraphs	t = -0.965	Equal	z = -1.610*	Equal
Number of Sentences	z = -1.707**	Equal	z = -4.188***	Different
Number of Words	z = -11.166***	Different	z = -4.722***	Different
<b>Panel B: Intelligibility Indexes</b>				
<b>Group 1: Connectives</b>				
All connectives	t = 18.727***	Different	z = -11.135***	Different
Additive connectives	t = 11.476***	Different	z = -5.156***	Different
Causal connectives	z = -6.632***	Different	z = -6.191***	Different
Logical connectives	z = -4.004***	Different	z = -3.613***	Different
Temporal connectives	t = 24.134***	Different	z = -11.281***	Different
<b>Group 2: Content Words and Personal Pronouns Incidence</b>				
Adjective incidence	z = -9.408***	Different	z = -0.855	Equal
Adverb incidence	t = -1.216	Equal	z = -3.749***	Different
Noun incidence	z = -8.276***	Different	z = -6.642***	Different
Pronoun incidence	z = -11.505***	Different	z = -11.505***	Different
Verb incidence	z = -10.434***	Different	t = -8.18***	Different
Personal pronouns incidence	z = -11.462***	Different	z = -10.376***	Different
<b>Group 3: Overlaps</b>				
Adjacent argument	z = -11.505***	Different	z = -11.372***	Different
Argument	z = -9.736***	Different	z = -10.701***	Different
Adjacent stem	z = -11.505***	Different	z = -11.504***	Different
Stem	z = -11.504***	Different	z = -11.505***	Different
Adjacent content word	z = -11.505***	Different	z = -11.505***	Different
<b>Group 4: Other Indexes</b>				
Mean hypernyms per verb	t = 110.443***	Different	z = -11.505***	Different
Type to token ratio	z = -10.415***	Different	z = -11.287***	Different
Modifiers per Noun Phrase	z = -11.506***	Different	t = 98.025***	Different
Noun Phrase Incidence	z = -9.099***	Different	t = 36.27***	Different
Words before Main Verb	t = 4.684***	Different	t = 4.14***	Different

Note. \*p-value < 0.1; \*\*p-value < 0.05; \*\*\*p-value < 0.01. The parametric tests are represented by the t, which is calculated according to the t-test. The nonparametric tests are represented by the z, which is calculated by the Wilcoxon test.

The means of Table 7 mostly present differences between English and Portuguese, but almost all variables have the same behavior when we look at the types of notes. Overall, the means and medians of the studied variables do not diverge much.

We verified that the Flesch index's means are higher in English, the non-native language, on both types of notes. When looking at the means test, one may conclude that they are, in fact, statistically different in a 5% significance level. These results are in accordance with Pasqualini et al. (2011), who analyzed short stories, and found that English texts were more readable than their versions in Portuguese.

“Incidence of temporal connectives” has an English mean over a 100% greater than the one in Portuguese for both notes. Nevertheless, when we tested if they were, on average, statistically equal

(Table 7), we rejected the hypothesis that all connectives variables were equal in both languages, in a 5% significance level. Overall, English (non-native language) had more connectives than Portuguese (native language), which is different from the findings of Pasqualini et al. (2011). As these authors say, connectives make the organization of a text, facilitating reading; thus, through our results, it is possible to say that English is easier to read because it presents a greater cohesion index. Nonetheless, this result could be due to what Lundholm et al. (2014) found about foreign firms to the United States with ADRs that write clearer texts to compensate for geographic distance and attract investors. Consequently, the analyzed firms could have somehow improved their non-native language with the purpose of attracting international investment.

We tested if both languages have the same scores for content words and personal pronouns, and the means presented some differences between English and Portuguese. In Financial Instruments, “Pronoun incidence” has means of Portuguese that are reasonably greater than the ones in English. Oppositely, for this type of note, “Personal pronouns incidence” has means in Portuguese that are a lot greater than those in English. All other variables look similar, however, when we tested if they were, on average, statistically equal (using t-test or Wilcoxon), we found that only to “Adverb incidence” in Financial Instruments and for “Adjective incidence” in Provisions, we could not reject the hypothesis that they were equal in a 5% significance level, thus, to all others, they can be considered different. It may imply that the sentence components are different in both languages, which seems reasonable once its construction is also different. For instance, in Portuguese, we do not use as many personal pronouns as in English, so, as Pasqualini et al. (2011) study results, we would expect a greater usage in the latter.

The overlaps have means that are reasonably greater in Portuguese than in English. This evidence is in accordance to what was found by Pasqualini et al. (2011). Their study found that indexes in English have smaller scores, which means a lower repetition of words, stems, and arguments, which makes the text more complex. All other variables are similar in both languages. To confirm if they were equal, we used a t-test or Wilcoxon. We found that all overlap variables were statistically different in both languages at a 5% significance level.

The last group is composed of all other variables not classified in any other group. Their means have differences between the two analyzed languages. The means of English are greater for: “Mean hypernyms per verb” and “Modifiers per Noun Phrase”. The “Modifiers per Noun Phrase” behavior was in accordance to Pasqualini et al. (2011), which suggests better Portuguese results than English. Similarly, even though the difference is not as high as the ones from the previous variables, English is also greater for “Words before Main Verb”, which means that, in this case, English requires a larger working memory load. To test if these differences were statistically significant, we performed a t-test or Wilcoxon test. We rejected the hypothesis that English and Portuguese are equal in all variables classified as ‘others’ at a 5% significance level.

### 4.3 Impact of OCPC 07

We performed an Analysis of Variance (ANOVA) to verify if the four analyzed years were statistically indistinguishable when it presented all the assumptions necessary for this test (normality and homogeneity of variance) on observations of all four years. When any of those two assumptions were not met, we performed a Friedman test. We tested the normality through a Kolmogorov-Smirnov test and tested the homogeneity of variances through Levene. We presented the results in Table 8.

At a 5% significance level, we could not reject the hypothesis that all years are equal in any readability indexes, except for “Mean syllables per word” (Provisions – English). Only after taking the year of 2012 from the Friedman analysis the hypothesis of equality was not rejected (sigma of 0.252). All other variables presented sigma > 0.05. It is evident that the readability indexes do not change much from one year to the other. These results conform with Lewis et al. (1986), who did not

find improvements in the Flesch index for “Operations review” extract from financial reports to employees.

**Table 8 – Mean test (from 2012 to 2015)**

Variables	Financial Instruments				Provisions			
	Portuguese		English		Portuguese		English	
	Normality	Sigma	Normality	Sigma	Normality	Sigma	Normality	Sigma
<b>Panel A: Readability Indexes</b>								
Flesch index	No	0.927	No	0.109	Yes	0.353	Yes	0.805
Mean sentences / paragraph	No	0.685	Yes	0.952	No	0.098	No	0.094
Mean syllables/word	Yes	0.874	No	0.383	No	0.114	No	0.014
Mean words/sentence	No	0.746	Yes	0.503	No	0.658	Yes	0.701
Number of Paragraphs	Yes	0.626	Yes	0.527	Yes	0.952	Yes	0.940
Number of Sentences	Yes	0.685	Yes	0.659	Yes	0.997	Yes	0.974
Number of Words	No	0.283	No	0.189	No	0.342	No	0.659
<b>Panel B: Intelligibility Indexes</b>								
<b>Group 1: Connectives</b>								
All connectives	Yes	0.776	Yes	0.557	No	0.293	No	0.007
Additive connectives	Yes	0.953	Yes	0.892	No	0.586	No	0.569
Causal connectives	No	0.094	No	0.491	No	0.810	Yes	0.763
Logical connectives	No	0.306	Yes	0.923	No	0.888	No	0.025
Temporal connectives	Yes	0.707	No	0.640	Yes	0.726	No	0.262
<b>Group 2: Content Words and Personal Pronouns Incidence</b>								
Adjective incidence	No	0.450	Yes	0.682	Yes	0.267	No	0.015
Adverb incidence	Yes	0.498	Yes	0.886	No	0.420	No	0.443
Noun incidence	No	0.065	Yes	0.908	Yes	0.974	Yes	0.732
Pronoun incidence	Yes	0.763	No	0.667	No	0.820	No	0.721
Verb incidence	No	0.165	No	0.186	Yes	0.956	Yes	0.629
Personal pronouns incidence	No	0.044	No	0.565	No	0.375	No	0.654
<b>Group 3: Overlaps</b>								
Adjacent argument	Yes	0.525	No	0.059	No	0.283	Yes	0.458
Argument	No	0.968	No	0.370	No	0.238	Yes	0.513
Adjacent stem	Yes	0.875	No	0.203	No	0.639	No	0.218
Stem	No	0.969	Yes	0.394	No	0.122	Yes	0.264
Adjacent content word	Yes	0.813	No	0.456	No	0.409	Yes	0.422
<b>Group 4: Other Indexes</b>								
Mean hypernyms/verb	Yes	0.789	Yes	0.915	No	0.867	Yes	0.628
Personal pronouns incidence	No	0.184	No	0.430	No	0.936	No	0.666
Type to token ratio	No	0.578	Yes	0.718	Yes	0.881	Yes	0.777
Modifiers / Noun Phrase	No	0.189	Yes	0.381	Yes	0.490	Yes	0.601
Noun Phrase Incidence	Yes	0.934	No	0.541	No	0.460	Yes	0.821

For Group 1 of the intelligibility indexes (Connectives), in a two-tailed test, we rejected the hypothesis that all years are equal, with a 95% confidence level, only for “Connectives incidence” and “Incidence of logical connectives”, both for Provisions notes in English. All other variables of group 1 have a sigma above 0.05. These two variables from 2012 to 2014 did not present a normal behavior. Thus, we performed a new Friedman test without 2015, and we found that the means of all other years can be considered equal, in a 5% significance level (sigma of 0.084 for the first and of 0.441 for the latter). It means that the year 2015 showed an improvement in the use of these writing tools.

For group 2 (Content Words and Personal Pronouns Incidence), we rejected the hypothesis that all years are equal, in two-tailed test and a 5% significance level, for “Adjective incidence” (Provisions – English) and for “Personal Pronouns incidence” (Financial Instruments – Portuguese). Thus, it is not possible to verify any increase or decrease in such indexes. However, it was not

expected for such variables to change much. What influences the understandability is how these types of words are organized rather than how much they appear in a text.

No variable of groups 3 (Overlaps) and 4 (Other Indexes) showed difference statistically significant at a 5% significance level in any language or type of note, which means that there is no variation from one year to the other, i.e., no improvement on the indexes after OCPC 07.

#### 4.4 Results Discussion

Our results showed significant differences between the intelligibility indexes in Portuguese and English. It was already expected, once as Pasqualini et al. (2011), they are languages which grammars work quite differently. Thus, these differences may be due to the translation problems mentioned by Evans et al. (2015) related to grammatical and language-specific lexical features. Another possible explanation is that considering that the native language of the writer was Portuguese, this may have been the reason why the intelligibility indexes were better for this language. It is in accordance with Evans et al. (2015), who say that functional translation, such as the one in Accounting, needs, among other translation abilities, a high degree of expertise in source and target languages, in the subject domain and knowledge of both source and target culture and conventions. Further investigation would be necessary on this, with opposite translations, from English to Portuguese, by native speakers of English.

Based on Lundholm et al. (2014), who found that foreign firms listed in the U.S. have better readability levels than U.S. firms to attract more American investment, we could expand this interpretation. We say that Brazilian firms showed better readability/intelligibility for Portuguese information because they might focus on local investors. New researches could test this assertion.

Previous research has discussed the differences between the concepts of readability and understandability (Smith & Taffler, 1992; Jones, 1997; Jones & Smith, 2014). Experiments (Soper & Dolphin Junior, 1964; Smith & Taffler, 1992) showed relations between understandability and readability, but little has been done to look for a different metric that could embody most of the understandability fundamentals. Considering intelligibility as a more complete metric for measuring understandability, and after comparing both proxies, we found evidence that readability and intelligibility metrics do measure different things. Overall, our results showed that the group with less readable information had better scores on a significant number of the intelligibility tests. To the other tests, the difference between groups was not statistically significant. Our results may imply that both metrics are incomplete or not enough to measure understandability. However, theoretically and considering all indexes that measure intelligibility, we may conclude differently. As Dreyer (1984), readability formulas do not measure word frequency, concept density, and abstraction level. Thus, as McNamara et al. (2002) believe, those formulas' quality is severely limited. The intelligibility scores, however, do measure those and other things (Graesser et al., 2004), which should make these metrics better at analyzing understandability. Experimental research could test the adequacy of intelligibility indexes for this purpose.

OCPC 07 required a better understandability of financial statements. However, our results allowed us to verify that the means of most readability and intelligibility indexes did not change after the OCPC 07 was issued. It could mean that firms did not improve their understandability, as requested by the guideline, or firms did change understandability, but neither metric measures understandability properly, which is unlikely. Our evidence is in accordance to Lewis et al. (1986), who could not find any relevant changes on the readability levels from 1977 to 1980 in any of the five methods they used. Similarly, Gomes et al. (2018) and Santos et al. (2019) also could not find any improvement in readability. What we found that was not verified by these studies is that the lack of improvement also applies to intelligibility indexes.

On the other hand, these results do not follow what Moreno and Casasola (2015) found, which was an improvement of readability over time, neither to what Soper and Dolphin Junior (1964) found:

a decrease of readability from 1948 to 1961. Even though the difference between results could be due to the method used, examining previous studies shows that the changes might take some time. Firms tend to copy and paste their notes from the previous periods (Santos et al., 2019). Therefore, studying a longer term might bring different conclusions. The referenced studies that had a null impact over time also analyzed only a 2 to 4 year-period. Moreno and Casasola (2015) might have found a different result because they analyzed reports of 7 to 8 decades.

The results are very similar for both types of notes, suggesting that understandability does not vary much throughout the Financial Statements.

Thus, we rejected hypothesis H1, i.e., high readability notes do not have high intelligibility. Similarly, we also rejected H3, which means there is no evidence of the improvement of readability and intelligibility with the OCPC 07. In contrast, we could not reject H2A and H2B, i.e., reporting language impacts on notes' readability and intelligibility.

## 5 CONCLUDING REMARKS

In this research, we analyzed the relation between readability and intelligibility. Our study is one of the few that analyses readability and applies intelligibility metrics to Financial Information.

We found that readability and intelligibility metrics measure different aspects of a text, which may indicate that readability formulas are not complete, and we should reconsider them as a measure for written financial information quality.

Moreover, our results indicated that Portuguese (the native language) and English (the non-native language) have different readability levels and intelligibility, which means that language is an important factor for text analysis. For instance, our results might indicate that the native language tends to be more readable and intelligible than the non-native language. Also, this finding increases the need for further investigation on understandability, once the results of scientific papers that analyzed reports in English may not be suitable parameters for analysis of reports in other languages.

At least regarding readability and intelligibility, after OCPC 07 was issued, there were no significant differences between both periods in these two metrics. It is evidence that firms do not change their reports' quality in such a short term, and we should consider new ways of enforcement methods if we want to apply them soon.

Our research's limitation is that the results may not be generalizable to other years or to other firms that do not have to present their information in English. Furthermore, the databases used by each software, Coh-Metrix, and Coh-Metrix-Port, are not the same, which can impact the results.

We suggest for future research: comparisons with other languages besides Portuguese and English; to analyze other accounting or financial documents, such as auditors' opinion, MD&A, and others, and test through an experiment how different levels of intelligibility impact understandability.

This research showed the relation between readability and intelligibility of notes to the financial statements, findings that may be useful for some groups. Firms or preparers may use the intelligibility indexes to analyze the ease of comprehension, improving their quality. We also presented an alternative measure for understandability that researchers may use instead of readability.

Furthermore, the evolution over the years showed that, at least in the writing quality, there is no evidence that firms have been improving their information after the OCPC 07. Thus, standard setters, regulators, and investors can use these results to demand better information quality from firms.

## REFERENCES

- Barth, M., Landsman, W., & Lang, M. (2008). International accounting standards and accounting quality. *Journal of Accounting Research*, 46(3), 467-498.
- Bloomfield, R. (2008). Discussion of "annual report readability, current earnings, and earnings persistence". *Journal of Accounting and Economics*, 45, 248– 252.
- Coh-Metrix-Port Website. (2017). *Coh-Metrix-Port tool*. Fonte: <http://143.107.183.175:22680/>

- Comitê de Pronunciamentos Contábeis – CPC. (2 de Dezembro de 2011). *Pronunciamento conceitual básico (R1): estrutura conceitual para elaboração e divulgação de relatório contábil-financeiro*. Fonte: Site do CPC: <http://www.cpc.org.br>
- Comitê de Pronunciamentos Contábeis – CPC. (26 de Setembro de 2014). *Orientação técnica OCPC 07 - evidenciação na divulgação dos relatórios contábil-financeiros de propósito geral*. Fonte: Site do CPC: <http://www.cpc.org.br/CPC/Documentos-Emitidos/Orientacoes/Orientacao?Id=95>
- Dreyer, L. G. (1984). Readability and responsibility. *Journal of Reading*, 27(4), 334-338.
- Evans, L., Baskerville, R., & Nara, K. (2015). Colliding worlds: issues relating to language translation in accounting and some lesson from other disciplines. *ABACUS*, 51(1), 1-36.
- EY. (2016). *SEC comments and trends: An analysis of current reporting issues*. U.S. Professional Practice Group.
- Finatto, M. J. (2011). Complexidade textual em artigos científicos: contribuições para o estudo do texto científico em português. *Organon*, 25(50), 30-45.
- Gomes, M., Ferreira, R., & Martins, V. (2018). O impacto da OCPC 07 sobre o tamanho e a legibilidade das notas explicativas de companhias brasileiras. *Revista Universo Contábil*, 14(2), 162-184.
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36(2), 193-202.
- Guay, W., Samuels, D., & Taylor, D. (2016). Guiding through the Fog: Financial statement complexity and voluntary disclosure. *Journal of Accounting and Economics*, 62, 234-269.
- Jones, M. J. (1997). Methodological themes: critical appraisal of the cloze procedure's use in the accounting domain. *Accounting, Auditing & Accountability Journal*, 10(1), 105-128.
- Jones, M., & Smith, M. (2014). Traditional and alternative methods of measuring the understandability of accounting narratives. *Accounting, Auditing and Accountability Journal*, 27(1), 183-208.
- Klare, G. R. (1974-1975). Assessing readability. *Reading Research Quarterly*, 10(1), 62-102.
- Kumar, G. (2014). Determinants of readability of financial reports of US-listed Asian companies. *Asian Journal of Finance & Accounting*, 6(2), 1-18.
- Lawrence, A. (2013). Individual investors and financial disclosure. *Journal of Accounting and Economics*, 56, 130-147.
- Leffa, V. J. (1996). *Aspectos da leitura: uma perspectiva psicolinguística*. Porto Alegre: Sagra: DC Luzatto.
- Lehavy, R., Li, F., & Merkley, K. (2011). The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The Accounting Review*, 86(3), 1087-1115.
- Lewis, N. R., Parker, L. D., Pound, G. D., & Sutcliffe, P. (1986). Accounting report readability: The use of readability techniques. *Accounting and Business Research*, 16(63), 199-213.
- Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, 45 (2), 221-247.
- Loughran, T., & McDonald, B. (2014). Measuring readability in financial disclosures. *The Journal of Finance*, 69(4), 1643-1671.
- Lundholm, R., Rogo, R., & Zhang, J. (2014). Restoring the tower of babel: how foreign firms communicate with U.S. Investors. *The Accounting Review*, 89(4), 1453-1485.
- Martins, T. B., Ghiraldelo, C. M., Nunes, M. G., & Oliveira Junior, O. N. (1996). Readability formulas applied to textbooks in Brazilian portuguese. *Notas do ICMSC-USP, Série Computação*, 28.
- McNamara, D. S., Louwerse, M. M., & Graesser, A. C. (2002). *Coh-Metrix: Automated cohesion and coherence scores to predict text readability and facilitate comprehension*. Memphis: Project funded by the University of Memphis.



- Merkley, K. J. (2014). Narrative disclosure and earnings performance: evidence from R&D disclosures. *The Accounting Review*, 89(2), 725-757.
- Moreno, A., & Casasola, A. (2015). A readability evolution of narratives in annual reports: a longitudinal study of two Spanish companies. *Journal of Business and Technical Communication*, 29, 1-34.
- Pasqualini, B., Scarton, C., & Finatto, M. (2011). Comparando avaliações de inteligibilidade textual entre originais e traduções de textos literários. *8th Brazilian Symposium in Information and Human Language Technology*. Sociedade Brasileira de Computação.
- Razik, T. A. (1969). A study of American newspaper readability. *The Journal of Communication*, 19, 317-324.
- Rennekamp, K. (2012). Processing fluence and investors' reactions to disclosure readability. *Journal of Accounting Research*, 50(5), 1319-1354.
- Rush, R. T. (1985). Assessing readability: formulas and alternatives. *The Reading Teacher*, 39(3), 274-283.
- Santos, E., Calixto, L., & Bispo, M. (2019). Impacto da OCPC 07 no enxugamento das notas explicativas das companhias brasileiras. *Revista Contabilidade e Finanças*, 30(79), 58-72.
- Scarton, C., Almeida, D., & Aluísio, S. (2010). Análise da inteligibilidade de textos via ferramentas de processamento de língua natural: adaptando as métricas do Coh-Metrix para o português. *Linguamática*, 2(1), 45-61.
- Schrivver, K. A. (1989). Evaluating Text quality: the continuum from text-focused to reader-focused methods. *IEEE Transactions on Professional Communication*, 32(4), 238-255.
- Smith, J., & Smith, N. (1971). Readability: a measure of the performance of the communication function of financial reporting. *The Accounting Review*, 552-561.
- Smith, M., & Taffler, R. (1992). Readability and understandability: different measures of the textual complexity of accounting narrative. *Accounting, Auditing & Accountability Journal*, 5(4), 84-98.
- Soper, F., & Dolphin Junior, R. (1964). Readability and corporate annual reports. *The Accounting Review*, 358-362.
- Worthington, J. S. (1977). The readability of footnotes to financial statements and how to improve them. *Journal of Reading*, 20(6), 469-478.