

**INVESTIGATING DIGITAL READING IN L2 TO CRITICIZE AND TO SUMMARIZE:
WORKING MEMORY CAPACITY AND READING PURPOSE INFLUENCING
STRATEGY USE**

Leonilda Procailo

Universidade Estadual do Centro-Oeste (Brazil)

leonilda@unicentro.br

Lêda Maria Braga Tomitch

Universidade Federal de Santa Catarina (Brazil)

leda@cce.ufsc.br

Received: 15/02/2020 - **Approved:** 13/04/2020

DOI: doi.org/10.17533/udea.lyl.n78a13

Abstract: The construction of meaning and inference generation considering low- and high-working memory span when readers read digital texts in L2 to criticize and to summarize were examined. Data collection included Reading Span Test, verbal protocol and post-reading tasks. The low-span group showed more misunderstandings in both purposes and a significant result of moderate correlation between reading time and reading to summarize. Both groups made more metacognitive comments when reading hypertexts to summarize and found strategies to cope with the demands of the nonlinear hypertext.

Key words: digital reading; working memory capacity, inference generation; reading purpose.

LA INVESTIGACIÓN DE LA LECTURA DIGITAL EN L2 PARA CRITICAR Y RESUMIR: LA INFLUENCIA DE LA CAPACIDAD DE LA MEMORIA DE TRABAJO Y DEL PROPÓSITO DE LECTURA EN EL USO DE ESTRATEGIAS

Resumen: Se investigó la construcción de significado y la generación de inferencias por lectores con memoria de trabajo de corto y largo alcance. Los participantes leyeron textos digitales en inglés para criticarlos y resumirlos. Luego, se sometieron a una prueba de amplitud lectora, protocolo verbal y tareas de poslectura. El grupo de corto alcance informó la falta de comprensión en ambas condiciones y un resultado significativo de correlación moderada entre el tiempo de lectura y leer para resumir. Ambos grupos produjeron más comentarios metacognitivos al leer hipertextos para resumir y encontraron estrategias para satisfacer las demandas del hipertexto no lineal.

Palabras clave: lectura digital; capacidad de memoria de trabajo; generación de inferencias; propósito de lectura.

1. Introduction

The availability of information in the digital space and the increase of time that people spend reading electronic media have brought about more freedom and choices to increment the experience, as well as have impacted sustained and engaged reading due to its navigational features (Liu, 2005). In this regard, it is paramount to investigate the recent change in the acts of reading, which involve Internet reading situations and other digital and hypertext environments. It is still an arena of much debate to consider that the relation between nonlinear digital text and readers is mediated by an environment full of tentative and uncertain interactions.

One relevant aspect in this discussion, which is the cognitive overhead generally approached in studies related to reading in the area, is working memory capacity (Zumbach & Mohraz, 2008). Working memory is an important construct that intends to explain how individuals vary in the way they cognitively process and store information during activities of high demands. Research on working memory capacity limitations regarding complex tasks has informed reading comprehension in both traditional printed text and nonlinear hypertext in the readers' mother tongue (L1) (Tomitch, 2003) as well as in second language (L2) (Alptekin & Erçetin, 2009; Fontanini & Tomitch, 2009).

Consistently supported by empirical studies, issues pertaining to the differences between linear and nonlinear hypertext reading still hold, even though digital texts have become more available. The role of coherence in inference generation and comprehension at local and global levels (Waniek, 2012), studies relating working memory capacity (WMC) and linear versus nonlinear reading (Fontanini & Tomitch, 2009; Lee & Tedder, 2003), and strategy use in hypertext reading (Cho, 2011) are all variables that point to some dissimilarities between linear texts and nonlinear hypertext reading.

Considering all those aspects, the main purpose of this study was to examine reading of linear digital and nonlinear hypertexts by low- and high-WMC readers of English as a L2. In this regard, the present study discusses part of a doctoral dissertation conducted with 30 participants from an English Language undergraduate course at a public university in Brazil. Through employment of a mixed-method approach in an exploratory manner, the low- and high-WMC readers' construction of meaning from linear and nonlinear hypertexts and the effects of reading purpose both the process and product of reading in different conditions were analyzed, in other words, reading to criticize/give opinion and to do a summary. In order to pursue the main objective of this study, the following research questions are proposed: Firstly, what types of inferences, as proposed by Linderholm and van den Broek (2002), are generated in verbal reports by low- and high-WMC readers in each condition- summarizing and criticizing/giving opinion in each text mode: linear digital and nonlinear hypertext reading? Secondly, how is the readers' WMC related to the ability to recall arguments in each text mode, linear digital and hypertext, as measured by the post-reading tasks in both conditions: reading to criticize/give opinion and reading to summarize? And thirdly, how is WMC related to reading time in each condition: reading to criticize/give opinion and reading to summarize?

2. Review of the Literature

2.1. Working Memory and Reading

Research focusing on the limitations of working memory stems from the need to look at the memory system and how it works as performing various activities at the same time.

Reading as a complex activity involving several factors, such as low- and high-level processes, schema activation, proposition encoding, interpretation, and inference generation is mediated by this limited capacity system called working memory (hence WM) (Tomitch, 2003).

Heitz, Unsworth, & Engle (2005) proposed the dissociation of WMC between controlled and automatic processing. In their model, controlled processing is subject to attentional control, thus effortful, capacity limitation. Attentional control is a voluntary, cognitive act which suppresses prepotent responses that are not relevant to the task at hand. Basically, in Heitz *et al.*'s model, WMC is the ability that reflects to what extent one can control attention in tasks that involve interference from competing information, at the same time one is able to maintain information available even in «...interference rich contexts as well as block irrelevant, distracting information» (p. 66).

Heitz, Unsworth, & Engle's (2005) view of WM seems to best explain the complex interplay between controlled and automatic recruitment of attention when it comes to understanding reading in complex environments such as the digital space, and this is the model to which this piece of research adheres to.

As aforementioned, the role WM plays in reading comprehension has gained prominence with research focusing on the limited capacity of this system as performing various activities at the same time. Research considering the relationship between WMC and the reader's main idea construction both in L1 and L2 and their strategy use have demonstrated that, when reading in L1, high-working memory readers used integrative strategies while the low-working memory ones relied on predictions, failing to update content after finishing reading, and thus, not being able to revise and/or test prior knowledge against textual information (see for example, Torres, 2003; Alptekin & Erçetin, 2009). Torres (2003) demonstrated how the main idea construction is positively correlated to WMC, in both L1 and L2 conditions. There are also differences in strategy use in both reading situations (L1 and L2) by both high- and low-span readers: comprehension monitoring was used by low-span readers in L2, while the same strategy was more present in high-span readers in L1, with the crucial difference that high-spans were able to take fix-up measures to overcome the problem.

With specific relevance to the present study is the research conducted by Alptekin & Erçetin (2009) as their experiment taxed both literal and inferential comprehension¹ in L2 reading in both low- and high-span readers. Important insights came out from their study as WMC can be considered to predict L2 reading performance. As the authors concluded, literal comprehension can be related to WMC only for readers with low L2 proficiency, as these readers tend to rely more on text-based processing. Consequently, inferential comprehension can be said to correlate well with WMC in the L2 reading situation.

2.2. New Forms of Literacy: Hypertext and Internet Reading

The availability of texts on the Internet not only has expanded the access to texts, but has also motivated the interest in writing and sharing information. As this tendency has grown, the interconnection of texts, documents and other media, such as sound, video and images define the nonlinearity of texts in the networked world. The so called *hypertext* is defined as «[...] a computer-mediated text in which highlighted words or titles enable readers to interactively determine the order and level of detail by serving as links to other excerpts or documents of supporting information» (Lee & Tedder, pp. 767-768, 2003). Rasmussen (2007), as presented by Afflerbach & Cho (2009), describes hypertexts as «made of blocks of text in the form of written text, pictures, video, and sound, chained together by electronic links» (p. 81). This view considers both the Internet reading situations as well as other hypertext environments, such as hypermedia, a combination of written text and other types of media.

In this context, the role of the reader has expanded as dramatically as have the functions of texts. Fox & Alexander (2009) classified the reader's activity, in a model called transitional extension, as «constructing meaning while connecting across texts; creating individual navigational path through links; considering author; responding interactively, building collaborative understanding» (p. 233). In this framework, the text is still static, but

1. For Alptekin & Erçetin (2009), literal understanding involves integrating micro- and macro-level propositions to build a cohesive textbase. Inferential comprehension is a coherent interpretation, integrating «the surface code, the textbase, and the reader's relevant knowledge [...] such that a mental model of the situation is formed in relation to what the author means» (p. 629).

it is also fluid, as it is presented in different modalities of, as put by the authors, «single or multiple linked propositional networks» (p. 233).

2.3. Inference Generation and Strategy Use

Inference is a process which guides any kind of interpretation in everyday life. From a plain conversation between two or more people, to trying to understand an accident scene, to reading, a great deal of information is inferred from what is not explicitly stated, be it in written or oral discourse. In reading, inference generation may reveal the degree of involvement that the reader employs in the task, which largely depends on goals/questions that they have when approaching the text, text type and genre (Roscioli & Tomitch, 2014). Second, goals/questions are totally idiosyncratic and are regarded as deliberate processes. According to Cho (2011) inferential reasoning is related to an important aspect of reading strategies use, which refers to deliberate effort to accomplish understanding of both explicit and implicit information conveyed in the text.

Investigations on kinds of inferences and measures have been used since about 1970 and a common challenge has been finding empirical measures to research the processes and the products of inferences. Therefore, measures that include verbal reports, post-reading tasks and other behavioral measures have been used in more recent studies in reading comprehension (Magliano, 1999).

Finally, the standards of coherence model propose that readers adjust their cognitive effort to the goal they want to achieve when reading (Vonk & Noordman, 1990). When reading to study, the reader may engage in a more text-based representation, generating more paraphrases and connecting inferences, employing effort to generate more benefits from the task: a better recall of the reading. Conversely, when reading for entertainment, the representation of the text will be more global, associating information from the text to one's own experiences, and making more evaluating comments and associations (Linderholm & van den Broek, 2002; van den Broek *et al.*, 2001). The reader's experience, his/her active and high competence brought to the task of reading in order to meet those standards do play a role in a coherent understanding of text (Cho & Afflerbach, 2017).

In order to pursue the main objective of the present study, the Linderholm & van den Broek's (2002) inference generation model was applied, classified as: associations, evaluative comments, elaborative inferences, predictive inferences, reinstatement inferences, metacognitive comments, paraphrases, text repetitions. Their study replicates previous findings by van den Broek *et al.*'s (2001) study with the same model and other studies with empirical support to the assumptions that readers adjust their cognitive processes and strategies to accomplish their reading goals. Specifically, when reading for study, they engage in a more effortful task, reading slowly and resorting to strategies that would result in learning goals (repeating the text and paraphrasing). Whereas, when reading for entertainment, readers do it in a faster pace, engaging in cognitive processes that allow for more associations, elaborations, and formulation of opinions.

Strategy use in reading has been researched since the 1970's referring to cognitive aspects of information processing and definitions of what strategies really mean are controversial since they are usually used interchangeably with skills in reading comprehension (Afflerbach, Pearson, & Paris, 2008). Paris, Lipson, & Wixson (1994) related strategic behavior in reading to intentions, choices, and efforts. In their view, a strategic behavior means an intention, a purpose on the part of the reader which begins even before the moment she/he meets the text. It means that one alternative is chosen over other possible ones. Afflerbach, Pearson, & Paris (2008) proposed that «reading skills are automatic actions that result in decoding and comprehension with speed, efficiency, and fluency and usually occur without awareness of the components or control involved» (p. 368).

As it can be seen, the difference between being a deliberate or automatic act lies in the distinction between strategy and skill. This distinction is not always clear-cut since a reader can make use of any of the two, both strategy or skill, at any time during reading. Being a skilled reader does not guarantee that one is always going to perform the task in a skillful manner.

2.4. Critical Reading and Summarization-Concepts

Regarding the first reading purpose proposed in this research (reading to criticize), it is important to ponder that inferential comprehension is viewed as a high-level process within the framework proposed by Gagné, Yekovich, & Yekovich (1993), as it demands more from the cognitive process. Text-based and knowledge-based information must be integrated in order to form the situation model of the text (van Dijk & Kintsch, 1983; Kintsch & van Dijk, 1978).

In summary, L2 reading classes have a great array of issues to tackle considering the power relationship exerted across various cultures and realities, including the student's own socio-cultural context. In this respect, hypertexts allow a more dynamic interaction among readers and texts, that require a more critical stance and an active response, by selecting navigational paths across links and considering the author (Fox & Alexander, 2009), as well as interpreting, evaluating aspects and content of texts and their purposes, which influence the best paths to achieve a goal (Afflerbach & Cho, 2009).

Regarding the second reading purpose of the present research, reading for summarizing, Tomitch (2012) discussed that summarizing is part of the day-to-day interaction with relatives and friends, which is part of human nature. Expository texts may vary the structure as they combine different types of argumentations, according to Tomitch, problem/solution, cause and effect, description, classification, definition, and others.

For the matter of summarizing texts in general, the following «macrorules» proposed by Brown and Day (1983), as summarized by Tomitch (2012, pp. 79-80) are being considered in the present study: deletion of trivial information, deletion of redundant information, superordination of lists, superordination of actions, selection of a topic sentence, and invention of a topic sentence.

3. Method

3.1. Participants

To develop the study proposed hereby, 30 participants² from intermediate to advanced level of English (6 for the pilot study and 24 for the actual study) volunteered after a brief presentation considering the steps: group-screening, participant-screening, and setting up section, as suggested by Cho (2011). Their age ranged from 18 to 46 years old, with an average of 23.33 years old (SD = 7.23). Participants were all undergraduate students of English language in 2016, at least in the end of the 4th semester of the English Literature course of a Federal University in Brazil. These students were considered proficient speakers and readers in English, as at this stage they are about to complete 50 % of the course that lasts 8 semesters, having a considerable experience reading expository texts, summarizing, and criticizing.

3.2. *The Stimulus Texts*

Two expository texts in L2 English in digital format, one linear and one hyperlinked³, with controversial issues were used to collect data. The texts were presented in a computer screen, font size 12, following Darroch, Goodman, Brewster & Gray's research (2005). Both texts were edited to have similar rhetorical structures, with arguments pro and con in similar numbers. Text 1, entitled *Should Animals Be Used for Scientific or Commercial Testing?* (662 words) and Text 2, entitled *Are Social Networking Sites Good for Our Society?* (661 words), which has four arguments in favor of social networking and four arguments against it.

Data collection was done in individual sessions. As the purpose of the present study was to focus on process and product of reading in English as a L2 and not on the performance in the target language, all instruments' instructions and tasks were in Portuguese.

2. The research proposal was approved by the institutional ethics committee (CEPSH-UFSC-CAAE: 54319416.0.0000.0121).

3. The model of hypertext adopted is closer in design to one of the models used in Klois, Segers, & Verhoeven's study (2013). It is a text without overview, maps or any other structure. It needs to be connected to the World Wide Web to be accessed, resembling an online Internet text, but which is bound to an electronic system. The only navigational method used was the contextual embedded linking within document in order to avoid any interference associated with different navigational options, such as menus, maps, and lists. The hyperlinked information was accessed by clicking on the expression *more*, highlighted in blue and underlined. The link-node hyperlinked text (seven, in total) preserves the sequence of the linear text and the transitional signals that relate the idea. Participants could only choose a link to follow, read the related node, check another link and go back to the primary text. Links embedded in nodes are limited to one supporting detail in each text.

The individual sessions were conducted by this researcher in a room equipped with a laptop computer and a software to capture on-screen activity, and a cell phone with a voice recorder application. Other materials such as instructions, questionnaires and tasks were provided in print format. Besides, the screen recording device named *oCam* aided in analyzing the readers' path through the process. Even though there was not a time limit previously allotted for the whole experiment, on average, participants took 90 minutes to conclude the experimental session, including the Reading Span Test (RST), the reading sessions, the post-reading tasks —also a demographic questionnaire and a self-evaluation questionnaire, which are not being analyzed hereby—. In the same way, even though reading time was not controlled, it was measured by examining the screen shots and the amount of time the participant remained silent, moved or not the mouse arrow to indicate where and when reading or rereading was taking place.

3.3. The Reading Span Test

The first experimental procedure in the present research was the RST in Portuguese (an adaptation from Daneman and Carpenter, (1980) devised by Tomitch, (2003) and all participants received training with six sentences, devised by this researcher. Even though participants are proficient in English, the use of the RST in L1 Portuguese is backed up by previous research, such as the one by Linck, Osthus, Koeth, & Bunting (2013). Torres (2003) also indicated that the RST in L1 can be positively correlated to reading comprehension in L1 as well as in L2. Once the span test measures both processing and storage, it is positively correlated with L2 knowledge and may affect storage of information, influencing the final score of the span test. In order to avoid L2 interference, we decided to use the RST in Portuguese.

The RST was composed of 60 unrelated sentences ranging from 13 to 17 words, divided in five sections, each section containing three trials. Sentences were presented in a Power Point format, 36-point black Calibri font, centered, halfway down the computer screen, in white background. Participants were instructed to read each sentence aloud to ensure processing, as proposed by Daneman & Carpenter (1980) and to memorize the last word of each sentence. At the end of each trial of three sentences, a pencil icon appeared on the

screen signaling the recall should begin. Soon after that, the participants wrote down as many words as they could remember, in the same order they were presented.

The memory span was calculated based on procedures used by Linderholm and van den Broek (2002), and also in the study by Friedman & Miyake (2004; 2005), and it is considered a lenient measure of score. The two studies by Friedman and Miyake highlight the total number of words recalled method. The scores from the RST ranged from 24.5 to 51.0. After having all scores, a medium split criterion was adopted and the scores were dichotomized into: low-spans (from 24.5 to 34.5); and high-spans (from 35.0 to 51.0).

3.4. Verbal Report

Verbal protocols to assess the cognitive processes of reading have been used as a mean to infer both the process and the product of reading (Pressley & Afflerbach, 1995). In reading, it refers to the process of self-observation and verbalization that a reader performs in relation to the cognitive activity of comprehension (Tomitch, 2007). In the present study, both concurrent and retrospective reports were applied. Participants were instructed to verbalize any thought that came to their minds during reading in Portuguese⁴. It could be related to any aspect of the task: vocabulary, ideas in the text, text structure, and difficulty related to the task. Since the reader was also supposed to submit a report at the end of a paragraph, it also involves the so called «retrospective verbalization» similar to Tomitch's (2003).

The verbal reports were audio recorded in mp3 format with the help of an application downloaded to a cell phone, resulting in a very high-quality audio. Verbal utterances of all participants in both conditions were transcribed in detail. Verbal protocols should be systematized so as to avoid the researcher's interference, as they should be as close to a natural process as possible (Samuels & Kamil, 2002). Taking this aspect into account and to refine the instrument and avoid wasting material, a pilot-study with a group of six students, whose participants belonged to 7th semester, was conducted. The pilot study also

4. Even though participants are fluent speakers of English, they were asked to speak in Portuguese as it was expected that they spontaneously reported their opinions, thoughts at any time without monitoring themselves, since their L1 is Portuguese.

helped defining the best approach, such as avoiding researcher interference, improving questionnaire and instructions, change the positions of links, among others.

3.5. Design

Both a between-subjects and a within-subjects design was used. The between-subject factor was WMC. Between- and within-subject factors are the following: text mode, reading purpose, and post-reading task.

3.6. Inference Categories Ratings

Two independent raters, who were blind to the conditions, plus the first author categorized 50 % of the verbal reports according to the inference category framework by Linderholm & van den Broek (2002). The remaining reports were categorized by the first author. Inter-rater reliability is 89 % between rater 1 and rater 2, and 90 % between rater 2 and rater 3.

Along with the categories proposed by Linderholm and van den Broek (2002), namely, associations, evaluative comments, elaborative inferences, predictive inferences, eeinstatement inferences, metacognitive comments, paraphrases, and text repetitions, two other categories were identified and added to the framework: summarizations (main idea of the paragraph) and misunderstandings.

3.7. Post-Reading Tasks

The post-reading task in the specific condition reading to criticize/give opinion aimed at verifying the ability of readers to both recall and evaluate the arguments used by different voices in the texts. The specific condition reading to summarize was intended to assess the readers' ability to recall the arguments in a summary of 200 to 250 words. In fact, the ability to summarize was of secondary importance to the objectives of the study. The task was intended to bring some meaning to the reading purpose, as discussed above in the reading to criticize condition.

3.8. Data Analysis-Variables and Measures

This research of deductive nature was conducted with the help of a mixed-method approach via the employment of different measures that could account for the processes and products of reading. This study employed the triangulation of quantitative and qualitative data analysis, an approach that has been extensively applied in the area of reading (Caldart, 2012; Tomitch, 2003) —to mention a few—.

The software environment for statistical analysis R (R Core Team, 2017) was used in the statistical data analysis alongside the graphical user interface RStudio (RStudio Team, 2016). One 2 x 2 x 2 Mixed ANOVA was run for each type of inference identified in research question 1. Means, standard deviations and *p*-values of possible main effects of each factor or interaction among variables will be summarized. For research questions 2 and 3, Spearman's correlation was run.

4. Results and Discussion

In general, 1 099 inferences were generated during reading both texts. The total inferences in each condition: Criticize: 555; Summarize: 544; Hypertext: 550; Linear: 549. Total inferences per WMC: Low-WMC: 550; High-WMC: 549.

4.1. Research Question 1 (RQ1)

In order to pursue the main objective of the study, RQ1 strives to investigate what types of inferences, as proposed by Linderholm & van den Broek (2002), are generated in verbal reports by low- and high-WMC readers in each condition-summarizing and criticizing/giving opinion in each text mode: linear digital and nonlinear hypertext reading.

The variables working memory capacity, text mode, and condition had no effects on inference generation at the 5 % level of significance. The lack of statistically significant results may be due to the reduced sample size ($n = 24$) which is split in four groups when conditions are looked at separately. However, some effects that approached significance at

the alpha level of $\alpha = .1$ were considered trends and are discussed in view of similar results in previous research. The decision to consider significance approach at the level of $\alpha = .1$ was intended to ensure sufficient power to the test and took into account previous studies that corroborate the assumption that there are differences in reading performance between low- and high-WM groups in similar circumstances to those addressed in the present study (Linderholm & van den Broek, 2002), and similar reading purposes that facilitate the generation of specific inferences (van den Broek *et al.*, 2001). In this respect, and based on discussion provided by Larson-Hall (2010) and Kline (2004), it was decided that a Type II error (in the case of increasing the alpha level) could contribute to the field in the sense that it points to an effect in the sample, which is backed up by previous research (for more information, see Naumann, Richter, Christmann, & Groeben, 2007, for similar alpha level use in reading research).

Post-hoc analyses with two paired samples *t*-tests were run to further investigate the effect of grouping the categories into textbase inferences (paraphrase, text repetition, and summarization) and situation model inferences (elaborative inferences, evaluative comments, and associations).

—***Effects of working memory capacity and condition on inference generation:*** There is no main effect of WMC on inference generation at the 5 % level of significance. However, at the alpha level of .1, high- and low WMC would differ in the amount and quality of inference generation. No main effect of reading purpose was evidenced in the generation of inferences in either condition: reading to criticize or to summarize.

—***Effects of text mode on inference generation:*** There is no main effect of text mode on inference generation at the 5 % level of significance. The main effect may be considered statistically significant at the $\alpha = .1$ level in both cases ($p = .080 < .1$; mean = 3.42; SD = 2.60 for elaborative inferences and $p = .096 < .1$; mean = 1.50; SD = 1.96 for associations). As it can be observed, in general, readers elaborated more when reading linear texts. Hypertext reading favored associations, though.

4.1.1. Types of Inferences and Their Relations to Conditions

—**Elaborative inferences:** No statistically significant main effect was observed. However, although the text mode would be significant at the alpha level of .1, it may be considered a trend. No statistically significant interaction was observed.

Moreover, there is no interaction of variables on the generation of elaborative inferences at the 5 % level of significance. It can be considered significant at the $\alpha = .1$ level, showing a trend, $F(1,20) = 3.39, p = .080 < .1$, (mean = 3.42; SD = 2.60). Both groups elaborated more when reading the linear version of texts. Even though the low-WMC group produced more elaborations than their counterparts when reading hypertexts, high-spans elaborated more in the linear version of the text. Elaborations could have been the predicted outcome taking the level of controversy of the topics in both stimulus texts into account.

—**Evaluative comments:** Neither statistically significant main effect nor interaction was observed. The observed difference between low- and high-WMC readers in the condition reading to criticize in the hypertext mode is not considered statistically significant $F(1,20) = 0.02, p = .880$). However, it shows a trend, which is consistent with results from the post-reading task in reading to criticize/give opinion condition. Low-span readers evaluated more when reading to criticize in the hypertext mode. Linear reading did not display differences between groups.

—**Summarizations:** Neither statistically significant main effect nor interaction was observed. Even though observation may point to differences between the number of summarizations generated by high- and low-WMC groups, the difference is not statistically significant $F(1,20) = 0.56, p = .463$).

—**Text Repetitions:** Neither statistically significant main effect nor interaction was observed. However, the interaction between WM group and condition would be significant at the alpha level of .1 ($F(1,20) = 4.07, p = 0.057 < .1$), (criticize: mean = 2.71; SD = 3.20; summarize: mean = 3.04; SD = 3.95). It may be considered a trend. The low-WMC readers repeated the text independently of reading condition. In closer inspection, high-WMC readers repeated the text more in the criticize than in the summarize condition.

—**Paraphrases:** Neither statistically significant main effect nor interaction was observed. Even though differences in the total number of paraphrases can be observed favoring the low-WMC group in the linear criticize condition, and the high-WMC group in

the summarize condition (hypertext), results are not considered statistically significant, $F(1,20) = 0.01, p = .935$.

—**Associations:** Neither statistically significant main effect was observed. However, text mode would be significant at the alpha level of .1. It may be considered a trend. No statistically significant interaction was observed. Both groups associated more when reading hypertexts. The low-WMC group outperformed the high one when reading hypertexts to criticize, reaching significance at .1 level, $F(1,20) = 3.06, p = .096 < .1$ (mean = 1.50; SD = 1.96).

—**Metacognitive comments:** Neither statistically significant main effect nor interaction was observed. However, the interaction between text mode and condition would be significant at the alpha level of .1 $F(1,20) = 3.39, p = 0.081 < .1$, (criticize: mean = 1.38; SD = 2.36; summarize: mean = 1.25; SD = 1.67). These interactions may be considered a trend.

—**Reinstatement inferences:** Neither statistically significant main effect was observed. No statistically significant interaction was observed $F(1,20) = 0.61, p = .445$. That is, there are no any effects of WMC on the generation of reinstatement inferences in neither text mode nor condition. The negative results will be further discussed in view of specificity of instruments utilized in the present study.

—**Predictive inferences:** Neither statistically significant main effect was observed. No statistically significant interaction was observed $F(1,20) = 1.49, p = .237$. That is, there are not any effects of WMC on the generation of predictive inferences in neither text mode nor condition. The negative results will be further discussed in view of the specificity of the instrument utilized in the present study.

—**Misunderstandings:** Neither statistically significant main effect nor interaction was observed at the 5 % level of significance. However, the WM Group would be significant at the alpha level of .1. Low-WMC readers revealed more misunderstandings than the high-WMC ones. If the alpha level of .1 is considered, $F(1,20) = 3.29, p = .085 < .1$ (mean = 1.32; SD = 1.49) in comparison to the high-WM group (mean = 0.65; SD = 1.09). It may show a trend. Those readers showed more misunderstandings than the high-WMC group when reading to criticize and to summarize in the linear mode.

The RQ1 under analysis strives to identify the types of inferences generated during reading and revealed through verbal reports of participants and how those inferences interact with other variables such as WMC, text mode and reading purpose.

On the other hand, it is important to say that discussion will be first conducted taking into account those categories which presented a statistically significant interaction at the alpha level of .1, namely, elaborative inferences, text repetitions, associations, metacognitive comments, and misunderstandings.

—**Elaborative inferences:** In closer examination, both groups elaborated more when reading the linear version of texts. However, high-WMC readers generated more elaborative inferences than the low-WMC ones in this text mode in both conditions. The low-WMC group elaborated more than the high-WMC one when reading to criticize in the hypertext mode, even though text mode did not show a statistically significant result.

Linear text reading may be less demanding in terms of cognitive load as compared to nonlinear hypertext, as suggested in studies by Lee & Tedder (2003, 2004), for example. However, the influence of hypertext also depends on its structure and the type of reading that is performed.

Considering that during linear text reading elaborative inferences were generated by both WM groups, one can conjecture that being the stimulus texts provocative, highly enhancing the reader to take a side, they elicited more opinions from both groups who felt at ease to draw on background knowledge. So much so that by looking at reading conditions separately, reading purpose did not significantly influence the generation of elaborations. Processes such as elaborations, associations and evaluations involve integrating background knowledge and opinions to the text's ideas, and are less demanding cognitively speaking (van den Broek *et al.*, 2001). On the other hand, elaborative inferences can also point to an overreliance on prior knowledge in detriment of text information ending up in deficiency in recalling certain details of the text and may hinder text comprehension (Tomitch, 2003). Both groups engaged in a type of processing in which prior knowledge was easily called upon. Especially the low-spans did so in the hypertext mode.

—**Text repetitions:** Reading to summarize did not yield significant results in both memory span groups as regards text repetitions. Low-WMC readers generated the same

average amount of text repetitions independently of reading condition. On the other hand, high-WMC readers generated more text repetitions in the criticize than in the summarize condition.

Reading to summarize may be seen as a more demanding reading task as compared to reading to criticize, as it requires more commitment to the internal structure of the text. Therefore, it was expected that readers produced more text repetitions in the summarize condition than in the criticize/give opinion condition. A possible explanation for the use of text repetitions by both groups independently of conditions could be the constraints derived from the verbal protocol *per se* in the way it was performed in the specific study. Having to read in English and report in Portuguese may have influenced the repetition of text by both WM groups as two important activities were taking place simultaneously: comprehension and translation in order to report. Repeating the text may have seemed faster and more at hand than trying to get the gist of ideas. Especially referring to low-spans, Oliveira (2016) found out that mental translation may have increased the demands of WM resources of low-WM spans as they read a text in English and recalled in L1 Portuguese, suggesting that translation can be detrimental to comprehension for those readers. It remains to be answered why the high-spans also engaged in such processing.

—**Associations:** Even though no statistically significant results were shown at $\alpha = .05$, both groups made more associations when reading hypertexts. The low-WMC group outperformed the high-WMC one when reading hypertexts to criticize. The generation of associations by low-WMC readers in hypertext reading confirms the trend when compared to the generation of elaborative inferences. Low-WMC readers also elaborated more when reading hypertexts to criticize. In the framework by Linderholm & van den Broek (2002) elaborative inferences are seen as necessary to comprehend the text, while associations are generally interpreted as non-relevant to the text's ideas. As discussed above, other researchers classify elaborations as necessary or non-necessary for comprehension, depending on the excessive use of prior knowledge so as to dismiss text information.

—**Metacognitive comments:** Reading to summarize in the hypertext mode generated more metacognitive comments by both memory span groups. However, the high-span group outperformed the low-span one. Both groups generated more metacognitive comments when reading linear texts to criticize than when reading linear texts to

summarize, but the high-spans generated more metacognitive comments in this condition. The low-WMC group generated the same average amount of metacognitive comments when reading to criticize in both text modes.

Metacognitive comments are viewed as demanding once they require verbalizations of thoughts, and this process would compete for attention and split resources in the effort to comprehend and reflect on comprehension at the same time (Linderholm & van den Broek, 2002). Both reading to summarize and hypertext reading may fall in this spectrum. Qualitative analysis on the kind of reading performed by both groups deserves discussion when looking at link access and the use of other strategies. Another explanation may be that skilled readers—who could also be high-spans—use cognitive strategies in a more successful way than less skilled readers (the low-spans). Caution is needed not to conclude that low-spans are not strategic. The difference between both groups under discussion may lie on the fact that strategies may vary according to more and less demanding processes to achieve learning goals.

—**Misunderstandings:** Low-WMC readers showed more misunderstandings than the high-WMC group when reading to criticize and to summarize in the linear mode.

Besides, being consistent with the trends hereby discussed so far and with previous research (Linderholm & van den Broek, 2002), these results can be discussed in view of what has been found so far in the literature regarding reading in L2. It is still an arena of discussion how WMC can affect reading comprehension in L2. Even though participants in the present study were regarded as proficient readers, interference of the variable L2 may have occurred. Again, based on Oliveira (2016), mentally translating the verbal utterances to Portuguese may have caused the low-WMC readers to overtax the cognitive system which in turn impaired comprehension.

4.1.2. Exploratory data analysis

To further analyze whether by grouping text-base and situation model inferences would account for differences in the processes generated and verbalized in verbal reports hereby analyzed, two paired samples *t*-tests were run with the sum of the elaborative inferences, evaluative comments and associations, considered here as situation model inferences (as the

first test's dependent variable), since they draw on previous knowledge, and the sum of summarizations, text repetitions and paraphrases, referred to here as text-based inferences, as the dependent variable of the second test. The boxplots showed no statistically significant differences in either case (Situation Model Inferences: $t(23) = 1.11, p = .278$; Text-Base Inferences: $t(23) = -0.32, p = .750$).

Even though they are not statistically significant, the boxplots in the criticize condition seem to display a more normal distribution than in the summarize condition. If this assumption is possible, then the scores in the post-reading criticize best reflect the reading process in this condition, an issue to be qualitatively discussed.

4.2. Research Question 2 (RQ2)

Looking at the RQ2, that is, how the reader's WMC is related to the ability to recall arguments in each text mode, linear digital and hypertext, as measured by the post-reading tasks in both conditions, reading to criticize/give opinion and reading to summarize, as previously reported, significant outliers were observed in the data. Since the classical and most well-known correlation coefficient, Pearson's r , is not robust to outliers, Spearman's ranks were analyzed instead. The correlations were not statistically significant. The lack of a well-defined pattern in the plots' lines corroborates the lack of relationship between the variables. Again, the small sample size may have contributed to the negative statistical results.

Even though significant correlations between WMC and scores were not identified in each condition, in the summarize condition scores in general were lower than in the criticize condition, approaching a floor effect. Only three participants scored 7.0 or higher in this condition in contrast to the results in the criticize condition. From the 10 top scores (7.0 or higher), six were obtained by low-spans in the criticize condition. By triangulating results from RQ1 and scores from the post-reading tasks, and qualitative analysis of post-reading results, it is possible to say that as low-spans produced more elaborative inferences, associations and evaluative comments in the criticize condition, it may have resulted in better recall in the post-reading task in this condition. Another more likely explanation for the current results is that low-WMC readers have some awareness of the fact that they have

limited resources and attempt to strike a balance between more and less demanding processes in an attempt to meet learning goals (Linderholm & van den Broek, 2002).

It is important to highlight that texts' characteristics may have played an important role in the results. Reading highly controversial topics to criticize may have enhanced the readers' search for specific information to look at in order to be able to express opinions, so that elaborating, associating and evaluating may have been choices to achieve the goal. Texts' characteristics may have benefited the generation of knowledge-based inferences by both groups.

According to Conway & Engle (1994), WMC plays an important role in the retrieval, since it relies on controlled search rather than on search which begins by automatic activation. It may explain both groups' poor results in the condition reading to summarize in the sense that readers benefit from clear instructions on what to consider while reading for a better recall. It may be even more important for low-spans. Therefore, being instructed to look at arguments before reading and being instructed to recall a specific number of arguments may have contributed to better scores in the criticize when compared to the summarize condition for the low-WMC group.

Another relevant aspect to account for when discussing the lower scores on the summarize condition by both WMC groups and when recalling is considered is that the results may have been affected by the instructions in the pre-reading, by limiting the task to summarization. The post-reading instruction in this condition did not specify the number of arguments to be included in the summary. Independently of WMC, clear instructions influence the outcomes in the products of reading, as proposed by McCrudden and Schraw (2007).

Having asked participants to just summarize the text may not have contributed to a complete account of the relations between this specific reading purpose and the post reading and for the perceived difference between low- and high-WMC readers demonstrated in previous research. Based on a previous study (Gil, Braten, Vidal-Abarca, & Stromso, 2010), it is reasonable to conclude that evaluating the quality of summaries and arguments may better demonstrate the difference between groups. It remains a question to future studies whether more guided post-reading summarizing tasks with clear criteria may highlight the difference between low- and high-spans and between reading purposes.

Moreover, reading to summarize without a more guided procedure, such as directing the participants to what focus attention to may have contributed to the general employment of misleading cognitive processes as summary quality was not measured. While summarizing, readers may have controlled the number of words (200-250 words) and the number of arguments might have been neglected or condensed in favor of a well-structured summary. For instance, two arguments in each text were similar, and could have been recalled as one. When summarizing, participants could have more closely focused on macrorules as those summarized by Tomitch (2012, pp. 79-80): deletion of trivial information; deletion of redundant information; superordination of lists; superordination of actions; selection of a topic sentence; invention of a topic sentence.

Having this possibility in mind, summaries were re-evaluated considering the use of the following criteria: a) selection of a topic sentence; invention of a topic sentence; recall of at least 4 main arguments out of 8; deletion of trivial or redundant information; it contains 200-250 words (7.0 to 10.00 points); b) selection of a topic sentence; invention of a topic sentence; recall of 2 to 3 main arguments out of 8; use of redundant information, repetition of arguments; it contains less than 200 words; it contains opinions and/or information that are not in the text (6.0 to 6.5 points). The summaries were considered to be average (7.0) or well-structured (8.0-10.0) if they contained a topic sentence that presented a general idea and recalled most or all arguments within the limits of 200-250 words and deleted the redundant information.

Still maintaining the pattern, most of those who recalled more arguments, also did better in the second criteria when the whole summary text is considered. However, grades in the second evaluation are higher. As it can be inferred from the comparison between different criteria, both WMC groups maintained the trend. However, scores in the second criteria seem to be more consistent with the total amount of text-base inferences generated in reading to summarize. That is, by changing criteria, the amount of text-based inferences (text repetition, summarizations, and paraphrases) generated by both WMC groups (low-spans: 114; high-spans: 124) in the condition reading to summarize did contribute to a concise summary. High-spans performed better in this condition than the low-spans as they produced more text-based inferences resulting in more learning goals. These results are more consistent with those in previous research (Linderholm & van den Broek, 2002; van

den Broek *et al.*, 2001), although the recall measure used in their study was number of idea units.

Validity of the second criteria applied in the present study specifically in RQ2 was not statistically tested and may be questioned as summaries were not submitted to raters who were blind to conditions. In any case, offline measures should be adapted so as to provide a safe ground of discussion and for a better account of cognitive processes occurred during online processes.

4.3. Research Question 3 (RQ3)

Referring to RQ3, that is, how WMC is related to reading time in each condition, reading to criticize/express opinion and reading to summarize, correlations between the scores from the RST and reading time were analyzed in both conditions, namely reading to criticize/give opinions and reading to summarize. As participants were encouraged to behave naturally during the reading process, it was predicted that links could not be accessed conforming to each individual's reading habit or need in the nonlinear hypertext. Therefore, the research question under analysis did not include text mode as an independent variable. In addition, observation of screen shots identified several participants who did not access all or some links. In total, participants spent 167 minutes and 22 seconds to read the hypertext and 193 minutes and 52 seconds to read the linear text, and this difference may occur, in part, due to some participants did not access some or all links, be it as a choice or as a result of disorientation.

Both groups took longer to read the texts in the summarize than in the criticize/give opinion condition. The low-WMC readers read more slowly than the high-WMC readers when reading to summarize ($p = .04 < .05$). Since the classical and most well-known correlation coefficient, Pearson's r , is not robust to outliers, Spearman's ranks were analyzed instead. While the correlation in the criticize condition approached significance at the .05 level ($p = .05$), the correlation in the summarize condition was just significant ($p = .04 < .05$).

Previous studies corroborate the assumption that low- and high-WMC readers emphasize different reading strategies when reading for different purposes. The study by

Linderholm & van den Broek (2002) is one of such examples. Based on that and on the results of reading time herein presented, low-WMC readers employ rereading strategies and high-WMC emphasizes comprehension monitoring. So, reading time was 40 % longer for the low-WMC group in the summarize condition.

With a slight difference, it is reasonable to assume that high-WMC readers did better than the low-spans in the post-reading summarize condition when quality of summaries is considered. Moreover, both WMC groups adjusted processing to reach learning goals, but the pattern of processing adopted by low-spans in the summarize condition affected the post-reading results in a negative way (Linderholm & van den Broek, 2002).

5. Final Remarks

It was the purpose of this study to analyze reading comprehension from the perspective of a cognitive process regarding working memory and its relation to inference generation and other cognitive processes in reading to criticize/give opinion and reading to summarize in two text modes: linear digital and nonlinear hypertext.

Specifically, the present study examined the reading process via online measures and reading products via post-reading tasks in order to answer the following research questions:

Overall, the analysis of the first research question showed a trend that is backed up by previous research and provided converging evidence to the assumption that low- and high-WMC readers do differ in the type of inferences generated and cognitive processes and strategies deployed during reading for different purposes. The results suggest that low-spans do find ways to cope with the difficulties of the task by resorting to strategies that require less from the cognitive system. On one hand, those readers elaborate, evaluate, and associate more when reading to criticize in the hypertext mode, suggesting that they select strategies to achieve a goal and/or use strategies according to the demands of the task. The difference between low- and high-span readers lies on the fact that high-spans may suppress cognitive processes that may not be relevant to learning and, therefore, generate more inferences that show commitment to the internal structure of the text, namely, summarizations, text repetitions, and paraphrases. Overall, both WM groups generated the same average amount of text repetitions across conditions by suggesting that somehow

repeating the text provided a safe ground for both groups. As mentioned before, although the participants of the present study were considered proficient speakers and readers of English, translating while reporting may have influenced the repetition of text due to an overload of the memory system engaged in getting the gist of the paragraph and having to translate it in order to report.

As for the metacognitive comments, results from the present study suggest that both WM spans generated more metacognitive comments when reading to summarize in the hypertext mode, but the high-spans outperformed the low ones. This suggests that the constraints of a reading purpose which seemed more demanding may have been augmented by the text mode, ending up in the decrease of such comments by low-spans.

Low-WMC readers generated more misunderstandings in both reading purposes, to summarize and to criticize in the linear mode. On one hand, this process may have been enhanced by the procedure of mentally translating the text into Portuguese during verbal report, which could have impacted low-spans in a negative way by overtaxing the cognitive system and constraining comprehension.

Taken as a whole, reading to summarize can be viewed as a more challenging reading situation, since it leads readers to monitor comprehension and employ the necessary effort to accomplish a goal. During hypertext reading, qualitative analysis shows that both groups took decisions that could benefit the process: link access shows more commitment to explore the entire text in this condition.

Reading to criticize stimulated the generation of associations by the low-spans. As discussed above, low-spans engaged in a strategic processing that would demand less from the cognitive system by drawing on background knowledge to learn from the texts. Even though being highly stimulated by the topics, high-spans seemed to suppress comments and thoughts that would not contribute much to the understanding, while low-spans seemed to feel freer to associate more in this condition.

From what is stated above and drawing on results from the second research question, that intended to establish relations between WMC and the post-reading recall scores, reading to summarize can be seen as a more demanding reading purpose, as compared to reading to criticize. In such a challenging reading situation, it is reasonable to conclude that pre- and post-reading instructions may enhance or hinder the type of information to be

recalled. Especially pre-reading clear instructions on the type of task that will follow the reading activity may diminish the effect that an artificial reading purpose, peculiar to experimental sessions, could exert on the results. Qualitative analysis of instructions given in both reading purposes under discussion suggest that once the purpose is given and not selected by participants, they should be clear enough so that information in the text is given enough attention so as to guarantee encoding in long-term memory and the consequent retrieval. These interpretations are in line with the views of WM by Heitz, Unsworth and Engle (2005), Conway & Engle (1994), and Daneman & Carpenter (1980).

In this sense, based on qualitative analysis of post-reading tasks, reading to summarize is best measured by evaluation of summaries rather than the number of arguments recalled, as it was first hypothesized. Number of arguments as a measure of memory retrieval seems to best suit the criticize condition, as the boxplots show a more normal distribution in this condition. It remains unclear, though, why low-spans performed better when reading to criticize than the high-spans as measured by recall if qualitative analysis is considered. Reading to criticize a text in which controversial topics pop up may not be challenging enough so as to tackle differences.

Significant results of moderate correlation between WMC and reading time in each reading purpose addressed in the third research question add support to the discussion pertaining to reading in demanding reading situations, such as reading to summarize. Low-spans took longer to read in this condition, probably by trying to find ways to cope with the difficulties of the task, such as rereading before reporting. Qualitative analysis showed that readers in this condition accessed most links when compared to reading to criticize when reading in the hypertext mode. Strategic readers do find a balance between costs and benefits in order to achieve a goal. Analysis of the quality of summaries points to the interpretation that both groups engaged in a more in-depth reading, taking longer to read in this condition. Low-span readers spent more time than the high-span ones. Both groups adjusted processing to achieve learning goals. While low-spans found strategies to cope with the constraints of the task, high-spans relied on text-based inferences, producing more summarizations, text repetitions, and paraphrases, ending up producing more concise and comprehensive summaries.

The discussion provided in the present research may help the reading practice in educational contexts in the sense that the teaching of reading should consider the students' social practice, be it in the academic context or in schools. While it is taken for granted that the Gen Z or the iGeneration is defined by their ability to cope with the contingencies of the information era and the virtual environment, social contexts have to be considered. Reading highly interactive texts for entertainment may not be a constraint for such a generation. Therefore, educational practice should consider the inclusion of digital reading in web-based reading space to ensure that students can slowly be literate in different text modes. Moreover, instruction could develop and support the use of strategies for comprehension and navigation skills, given the changing nature of reading with the advent of computerized reading and the Internet availability of texts.

6. Limitations of the Present Study and Suggestions for Future Research

A few limitations to this study deserve discussion:

—**Number of participants:** The lack of statistical power in the first two research questions may, in part, be attributed to the small number of participants.

—**Criteria for splitting groups:** The criterion adopted in the present study was the medium split, ending up in two groups. By having a larger sample, the upper- and lower-third criterion should be adapted, so as to have a clear difference between groups.

Bibliographical References

1. Afflerbach, P. & Cho, B. (2009). Identifying and describing constructively responsive comprehension strategies in new and traditional forms of reading. In S. E. Israel & G. G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 69-90). Routledge, New York.
2. Afflerbach, P., Pearson, P. D., & Paris, S. G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher*, 61(5), 364–373.
Retrieved from <http://www.academia.edu/3141569>

3. Afflerbach, P. (2000). Verbal reports and protocol analysis. In M. L. Kamil, P. B. Mosenthal, D. Pearson, & R. Ban (Eds.), *Handbook of reading research -Vol. III* (pp. 163-179). New Jersey: LEA. Retrieved from <http://www.questia.com>
4. Alexander, P. A. & Fox, E. (2004). A historical perspective on reading research and practice. *Theoretical models of reading* (pp. 33-68). IRA. Retrieved from <https://www.researchgate.net/publication/255654818>
5. Alptekin, C. & Erçetin, G. (2009). Assessing the relationship of working memory to L2 reading: Does the nature of comprehension process and reading span task make a difference? *System*, 37, 627-639. DOI: 10.1016/j.system.2009.09.007
6. Caldart, D. (2012). *The effect of genre expectation on EFL Brazilian students' inference generation and reading comprehension*. (Master thesis). Universidade Federal de Santa Catarina, Florianópolis.
7. Cho, B.-Y. (2011). *Adolescents' constructively responsive reading strategy use in a critical internet reading task* (Doctoral dissertation). University of Maryland College Park, Maryland. Retrieved from <https://drum.lib.umd.edu/handle/1903/11879>
8. Cho, B.-Y. & Afflerbach, P. (2017). An evolving perspective of constructively responsive reading comprehension strategies in multilayered digital text environments. In S. E. Israel (Ed.), *Handbook of research on reading comprehension*, (2nd Edition) (pp. 109-134). Guilford Publications. Retrieved from <https://www.researchgate.net/publication/317021664>
9. Conway, A. R. & Engle, R. W. (1994). Working memory and retrieval: a resource-dependent inhibition model. *J Exp Psychol Gen.*, 123(4), 354-73.
10. Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
11. Darroch, I., Goodman, J., Brewster, S., & Gray, P. (2005). The effect of age and font size on reading text on handheld computers, In M. F. Costabile & F. Patemò (Eds), *Human-computer interaction-INTERACT 2005: Proceedings of the IFIP TC13 International Conference* (Rome, Italy), *Lecture Notes in Computer Science*, number 3585. Berlin: Springer-Verlag, (pp. 253-266). Retrieved from https://link.springer.com/chapter/10.1007/11555261_23

12. Erceg-Hurn, D. M. & Mirosevich, V. M. (2008). Modern Robust Statistical Methods: An easy way to maximize the accuracy and power of your research. *American Psychologist*, 63(7), 591-601.
13. Fontanini, I. & Tomitch, L. M. B. (2009). Working memory capacity and L2 university students' comprehension of linear texts and hypertexts. *IJES*, 9(2), 1-18. Retrieved from <https://www.researchgate.net/publication/40910165>
14. Fox, E. & Alexander, P. A. (2009). Text comprehension: a retrospective, perspective, and prospective. In S. E. Israel & G. G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 227-239). Routledge, New York
15. Friedman, N. P. & Miyake, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods*, 37(4), 581-590. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/16629290>
16. Friedman, N. P. & Miyake, A. (2004) The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language*, 51, 136-158. Retrieved from <https://doi.org/10.1016/j.jml.2004.03.008>
17. Gagné, E. D, Yekovich, C. W., & Yekovich, F. R. (1993). Reading. In *The cognitive psychology of school learning* (pp. 257-312). New York: Harper Collins College Publishers.
18. Gil, L., Braten, I., Vidal-Abarca, E., & Stromso, H. I. (2010). Understanding and integrating multiple science texts: summary tasks are sometimes better than argument tasks. *Reading Psychology*, 31, 30-68. DOI: 10.1080/02702710902733600
19. Heitz, R. P, Unsworth, N., & Engle, R. W. (2005). Working memory capacity, attention control, and fluid intelligence. In O. Wilhelm, R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 61-77). Thousand Oaks, CA: Sage Publications, Inc. DOI: 10.1016/j.cogpsych.2014.01.003
20. Kintsch, W. & van Dijk. T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363-394.
21. Kline, R. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, D. C.: American Psychological Association.
22. Klois, S. S., Segers, E., & Verhoeven, L. (2013). How hypertext fosters children's knowledge acquisition: The roles of text structure and graphical overview. *Computers in*

- Humans Behavior*, 29, 2047-2057. Retrieved from <http://dx.doi.org/10.1016/j.chb.2013.03.013>
23. Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York and London: Routledge. Retrieved from <http://cw.routledge.com/textbooks/9780805861853/R/full-version.pdf>
24. Lee, M. & Tedder, M. (2003). The effects of three different computer texts on readers' recall: based on working memory capacity. *Computers in Human Behavior*, 19, 767-783. Retrieved from <https://www.researchgate.net/publication/221998003>
25. Lee, M. & Tedder, M. (2004). Introducing expanding hypertext based on working memory capacity and the feeling of disorientation: tailored communication through effective hypertext design. *Journal of Educational Computing Research*, 30, 171-195. Retrieved from <https://www.researchgate.net/publication/250144673>
26. Linck, J. A., Osthus P., Koeth, J. T., & Bunting, M. F. (2013). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin and Review*. DOI 10.3758/s13423-013-0565-2
27. Linderholm, T. & van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology*, 94(4), 778-784. DOI: 10.1037//0022-0663.94.4.778
28. Liu, Z. (2005). Reading behavior in the digital environment: changes in reading behavior over the past ten years. *Journal of Documentation*, 61(6), 700-712. Retrieved from <http://www.emeraldinsight.com/researchregister>
29. Magliano, J. P. (1999). Revealing inference processes during text comprehension. In S.R. Goldman, A.C. Graesser, P., & van den Broek (Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 55-75). Mahwah, NJ: LEA. Retrieved from <https://www.questia.com/read/27719794/narrative-comprehension-causality-and-coherence>
30. McCrudden, M. T. & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review*, 19, 113-139. DOI: 10.1007/s10648-006-9010-7
31. Naumann, J., Richter T., Christmann, U., & Groeben, H. N. (2008). Working memory capacity and reading skill moderate the effectiveness of strategy training in learning from

hypertext. *Learning and Individual Differences*, 18(2), 197-213. DOI: 10.1037/0022-0663.99.4.791

32. Oliveira, D. A. (2016). *Working memory capacity and mental translation in EFL reading comprehension* (Master thesis). Universidade Federal de Santa Catarina, Florianópolis.

33. Paris, S. C., Lipson, M. Y., & Wixson, K. K. (1994). Becoming a strategic reader. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (pp. 788-810). Newark: International Reading Association.

34. Pressley, M. & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. New Jersey: LEA. Retrieved from <http://www.questia.com>

35. R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>

36. Roscioli, D. C. & Tomitch, L. M. B. (2014). The effect of genre expectation on EFL Brazilian students' inference generation and reading comprehension. *Revista do GEL*, São Paulo, 11(1), 73-109. Retrieved from <https://revistadogel.gel.org.br/rg/article/view/216>

37. RStudio Team (2016). *RStudio: Integrated Development for R*. RStudio, Inc., Boston, MA. Retrieved from <http://www.rstudio.com/>

38. Samuels, S. J. & Kamil, M. L. (2002). Models of the reading process. In R. Ban, M. Kamil, & P. Mosenthal, (Eds), *Handbook of reading research. Vol. 1* (pp. 185-224). New Jersey: Lawrence Erlbaum Associates, 2002. Retrieved from <http://www.questia.com>

39. Tomitch, L. M. B. (1995). *Reading: text organization perception and working memory capacity* (Doctoral dissertation). Universidade Federal de Santa Catarina, Florianópolis.

40. Tomitch, L. M. B. (2003). *Reading: text organization perception and working memory capacity*. Florianópolis: UFSC, DLLE.

41. Tomitch, L.M.B. (2007). Desvelando o processo de compreensão leitora: protocolos verbais na pesquisa em leitura. *Signo*, 32(53), p.42-53. Retrieved from <http://dx.doi.org/10.17058/signo.v32i53.244>

42. Tomitch, L. M. B. (2012). *Produção textual acadêmica*. Universidade Federal de Santa Catarina, Florianópolis.

43. Torres, A. C. G. (2003). *Working memory capacity and reader's performance on main idea construction in L1 and L2* (Doctoral dissertation). Universidade Federal de Santa Catarina, Florianópolis.
44. van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, 29(8), 1081-1087. Retrieved from <https://link.springer.com/content/pdf/10.3758%2FBF03206376.pdf>
45. van Dijk, T. & Kintsch, W. (1983). *Strategies for discourse comprehension*. New York: Academic Press.
46. Vonk, M. & Noordman, L. G.M. (1990). On the control of inferences in text understanding. In D.A. Balota, G.B. Flôres d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 447-464). LEA, NJ. Retrieved from <https://www.questia.com/read/59264892>
47. Waniek, J. (2012). How information organisation affects users' representation of hypertext structure and content. *Behaviour & Information Technology*, 31(2), 143-154. Retrieved from <http://dx.doi.org/10.1080/01449290903544652>
48. Zumbach, J. & Mohraz, M. (2008). Cognitive load in hypermedia reading comprehension: Influence of text type and linearity. *Computers in Human Behavior*, 24(3), 875-887. DOI: 10.1016/j.chb.2007.02.015