Language Teaching Research Quarterly

2022, Vol. 31, 136–158

Does Working Memory Capacity Predict Literal and Inferential Comprehension of Bilinguals' Digital Reading in a Multitasking Setting?

Bruno de Azevedo[1,2*], Davi Alves Oliveira[3], Ingrid Finger[4], Lêda Maria Braga Tomitch[1]

[1]Universidade Federal de Santa Catarina, Florianópolis, Brazil

[2]Instituto Federal de Santa Catarina, São Lourenço do Oeste, Brazil

[3]Universidade do Estado da Bahia, Jacobina, Bahia, Brazil

[4]Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

Received *25 May 2022* ***Accepted*** *05 September 2022*

Abstract

The ubiquity of multitasking has led researchers to investigate its potential costs for reading and learning (Clinton-Lisell, 2021). While some studies have not shown detrimental effects of multitasking for reading comprehension (Bowman et al., 2010; Cho et al., 2015; Pashler et al., 2013), one particular study has found a benefit of multitasking (Tran et al., 2013). These results, nevertheless, do not converge with the findings of recent meta-analyses, which have suggested both a negative effect of multitasking for reading comprehension (Clinton-Lisell, 2021), as well as the disruptive effects of listening to lyrical music while reading for comprehension (Vasilev et al., 2018). Previous research seems to converge with the theories of how working memory copes with the complexity of reading as a process, since several subprocesses must be orchestrated so that the ultimate goal of reading – the construction of a mental representation – is fully achieved (Tomitch, 2020). In addition to that, no previous study has investigated reading as a multilevel construct in which both literal and inferential comprehension (Alptekin & Erçetin, 2010; Kintsch, 1998) is assessed in a multitasking setting. With that in mind, we investigated whether working memory capacity, measured by the Self-Administrable Reading Span Test (Oliveira et al., 2021), predicts proficient bilinguals' performance in literal and inferential comprehension, by means of comprehension questions (Pearson & Johnson, 1978) and reading times, under a multitasking setting in two conditions – listening to lyrical music (experimental) as opposed to

bruno.azevedo@ifsc.edu.br doi:10.32038/ltrq.2022.31.10

listening to non-lyrical music (control). Multiple linear regression analyses revealed that working memory capacity significantly predicted inferential, but not literal comprehension nor reading times, and only when participants were listening to lyrical music. Results are discussed both in terms of the effects of multitasking on reading comprehension as well as the role of working memory in language comprehension.

Keywords: *Working Memory, Multitasking, Literal Comprehension, Inferential Comprehension, Digital Reading*

Introduction

The joint performance of two or more tasks simultaneously, henceforth multitasking, is a ubiquitous phenomenon (Clinton-Lisell, 2021; Salvucci & Taatgen, 2011). While some consider it as an ability in the 21st century, others conceive multitasking as a distraction (Aagaard, 2019), a debate which has gathered researchers from several fields investigating the potential consequences of multitasking for learning (May & Elder, 2018; Vasilev et al., 2018).

In reading comprehension, a recent meta-analytic review revealed the detrimental effect of multitasking on comprehension (Clinton-Lisell, 2021). Human cognitive resources are limited and thus multitasking might be costly for working memory (WM), which is the system employed in carrying out complex cognitive tasks such as reading (Baddeley, 2012). Daneman and Carpenter (1980) argue that, during reading, WM is highly deployed for storing "pragmatic, semantic, and syntactic information from preceding text" (p. 450) and it is what underlies syntactic parsing, inference generation, and subsequent integration of text parts. Similarly, to be effective as a reader, one must be able to "store the theme of the text, the representation of the situation to which it refers, the major propositions from preceding sentences, and a running, multilevel representation of the sentence that is currently being read" (Just & Carpenter, 1992, p. 122). Moreover, WM is resource-limited for executing these processes (Just & Carpenter, 1992), and individuals vary in their capacity to hold and manipulate information during reading (henceforth WMC - Working Memory Capacity) (Daneman & Carpenter, 1980). Thus, WMC is a good predictor of reading comprehension (Daneman & Merikle, 1996; Linck et al., 2014; Tomitch, 2020). However, little is known regarding WMC and multitasking (Pollard & Courage, 2017).

Based on these considerations, this study has investigated the effects of WMC in multitasking where reading is a primary task. In the present study, the representational architecture of reading comprehension is devised in terms of levels (Gagné et al., 1993; Kintsch, 1998), in which *literal comprehension* (lower-level) entails word decoding, lexical access and syntactic parsing processes, while *inferential comprehension* (higher-level) entails bridging and elaborative inferences, as well as main idea construction (Gagné et al., 1993). The different levels of comprehension pose distinct WM demands, as in proficient reading, literal comprehension might involve fewer working memory resources when compared to inferential comprehension (Alptekin & Erçetin, 2011; Just & Carpenter, 1992). For that reason, one of the hypotheses put forward in this study is that multitasking might be more detrimental to inferential comprehension, due to WM demands.

Another possible source of WM burden might be reading medium (DeStefano & LeFevre, 2007; Salmerón et al., 2018). Reading a digital text, most specifically a hypertext – a non-linear text with links that provide access to other pages which might contain other texts, videos, ads, among many others (Bråten et al., 2020) – might require more WMC given the array of processes and skills involved in hypertext reading (DeStefano & LeFevre, 2007; Salmerón et al., 2018). Considering the pervasiveness of hypertext reading due to the widespread use of digital devices in the last decades, this study investigated multitasking in hypertext reading.

The role of working memory in multitasking has been connected to Executive Functions (EF), "a set of general-purpose control mechanisms (…) that regulate the dynamics of human cognition and action" (Miyake & Friedman, 2012, p. 8). For successful multitasking, people need to *inhibit* dominant input in order to focus; *switch* between tasks; and *update* working memory representations. With this in mind, some researchers claim that knowing and using two or more languages – being bilingual (Bhatia et al., 2012), could be equated to a multitasking situation, and consequently, bilinguals would be better multitaskers (Poarch & Bialystok, 2015; Sörman et al., 2017). Under the recognition that the two languages of the bilingual are always active, even when the intention is to use only one (Kroll et al., 2006), processes such as *inhibiting* one language while using another and language *shifting* are assumed here to involve the same cognitive system used in multitasking.

Despite extensive research attempting to support the hypothesis that bilingualism bolsters executive functions (see Paap, 2019 for a review), the debate is full of controversies (Morton & Harper, 2007; Paap, 2019; Poarch & Krott, 2019), which might be attributed to the complexity of bilingualism as a construct (Backer & Bortfeld, 2021; DeLuca et al., 2019; Paap, 2019). To be more specific, DeLuca et al. (2019) explain that "bilingualism has been routinely operationalized as a categorical variable (bilingual/monolingual), whereas it is a complex and dynamic experience with a number of potentially deterministic factors" (p. 1), and consequently, should be taken into account by researchers (Scholl et al., 2021). Aspects related to the language background of bilinguals influence the categorization of bilingual participants (Scholl et al., 2017, 2021). Thus, bilinguals' language background and self-rated proficiency were considered here in order to explore any relationship between multitasking during reading in adult proficient bilinguals.

Based on these considerations, the present study aims at investigating whether working memory capacity, measured by the Self-Administrable Reading Span Test (Oliveira et al., 2021), predicts literal comprehension, inferential comprehension and reading times of a hypertext in proficient bilinguals in a multitasking setting in two conditions – listening to lyrical music (experimental) as opposed to listening to non-lyrical music (control).

Review of Literature

Working Memory in Reading

Working memory (WM) is highly deployed in complex cognitive tasks, such as reading, listening, learning and reasoning (Baddeley et al., 2021; Cowan et al., 2021; Sparks, 2019, 2021), and one of the sources of differences in carrying out such tasks comes from the capacity individuals have to hold and manipulate information during processing (Daneman & Carpenter, 1980; Tomitch, 2020). The nature of these individual differences in WMC is believed to have roots in the trade-off between storage and processing, which compete for resources, since a task that demands heavy processing lacks resources for storage (Daneman & Carpenter, 1980).

Alternatively, storage and processing in working memory are mediated by *activation,* and WMC is determined by the upper limit of activation WM can hold to support either storage or processing (Just & Carpenter, 1992). Individual differences in WMC can be explained both by *total-capacity* and *processing efficiency.* The former posits that task demands determine the amount of activation in WM, that is, differences in performance among readers is less apparent in easy comprehension tasks, and more evident in more difficult (or demanding) tasks (Just & Carpenter, 1992). Just and Carpenter contend that there is an 'optimal' level of task difficulty so that differences can be seen in the performance of lower- and higher-span readers, meaning that if the task is extremely difficult a floor effect may be observed for both groups (see, for example Tomitch, 2003). The *processing efficiency* determines resource allocation when the demand exceeds the supply, for instance, higher-level processes, considered more demanding than lower-level processes, might be compromised due to inappropriate supply of resources (Just & Carpenter, 1992). In other words, *processing efficiency* is connected with the level of 'expertise' in a given task, for instance, in additional language reading, lower-level comprehension processes must be processed automatically (thus, efficiently) so that there are enough cognitive resources left for higher-level comprehension processes.

In order to assess the role of individual differences in WMC, Daneman and Carpenter (1980) devised a test which places a burden on both storage and processing resources of WM. The Reading Span Test (RST) consists of having participants read a set of sentences for subsequent recall of the last word of each sentence, based on the assumption that sentence-final word recall should be easier for higher span readers given their resource availability (Daneman & Carpenter, 1980). The RST has been consolidated as a reliable research tool for assessing the relationship between WMC and reading (see Leeser & Herman, 2022; Tomitch, 2020 for reviews).

Despite WMC being positively related to second language (L2) reading outcomes (Linck et al., 2014; Sparks et al., 2019; Tomitch, 2020), some variables might have influenced the relationship between WM and L2 reading comprehension. Put differently, "methodological features of the RST may be moderator variables, variables that influence the strength of the relationship between the primary variables of interest" (Shin, 2020, p. 878). WM Task Content (e.g. verbal, nonverbal, or combined), for instance, may moderate the relationship between WM and L2 reading comprehension (In'nami et al., 2021).

Verbal span tasks are somehow connected to the *domain-specific* view of WM, while nonverbal span tasks are related to the *domain-general* view (In'nami et al., 2021; Linck et al., 2014). Domain-specific span tasks, such as the RST (Daneman & Carpenter, 1980), are believed to be highly correlated to language-related outcomes, such as reading (Linck et al., 2014) and listening comprehension (Sparks, 2019) given their higher sensitivity to differences in language-related processes (Wen, 2016). Domain-general span tasks, on the other hand, are believed to tap general WM resources regardless of task content (verbal or nonverbal) (Engle, 2018; Wen, 2016).

The *domain-specific* view holds a structural perspective of WM (Wen & Li, 2019), comprising a modular model where each component is held responsible for a *domain-specific* process (Baddeley, 2017). Baddeley's Multicomponent Model of Working Memory is composed of components for temporarily storing visual and phonological information (the sketchpad and the loop, respectively), a component for integrating information in the environment and long-term memory (the episodic buffer), and the Central Executive, which coordinates all these components (Baddeley, 2017; Baddeley et al., 2021). As opposed to that, the *domain-general* approach features "the attention-oriented and executive control views of WM" (Wen & Li, 2019, p. 369), advocated by several researchers (Cowan et al., 2021; Just & Carpenter, 1992; Mashburn et al., 2021 to mention a few).

In their Embedded-Processes Approach to WM, Cowan et al. (2021) highlight the role of control and attention in "shuttling items into and out of the focus of attention, allowing binding and conceptual formation to take place" (p. 45). Both control and attention are key features of a multitasking setting, in which individuals must *inhibit* dominant input in order to focus; *switch* between tasks; and *update* WM representations (Miyake & Friedman, 2012). Another *domain-general* approach views WM as the "ability to maintain information in the maelstrom of divergent thought" (Engle, 2018, p. 192). Most recently, Engle's group has argued that WMC is situated in a broader framework of complex cognition which is controlled by an executive attention system, in which attention is unitary in the sense that it "cannot be divided but can be switched back and forth between tasks quickly" (Mashburn et al., 2021, p. 176).

Nevertheless, some scholars seem to converge towards efforts of integrating WM theories, especially those concerned with the practical applications of WM models with general cognition (Wen & Li, 2019), considering that a single view of WM cannot account for the complexity of cognitive and language processing (Logie et al., 2021; Wen, 2016; Wen & Schwieter, 2022). More specifically, phonological and visual information maintained in WM are deemed to be *domain-specific* (in Baddeley's model), while attentional/executive control processes are considered *domain-general*, thus, Wen and Schwieter (2022) suggest that we should consider WM "as a multicomponent system that consists of both domain-specific storage buffers and domain-general executive control functions (Conway et al., 2005). Therefore, neither a completely domain-general nor a completely domain-specific view of working memory holds" (p. 913).

The Cognitive Architecture of Comprehension

The construction of a coherent mental representation is the outcome of comprehension (Kintsch, 1998, 2013) and entails the orchestration of both lower- and higher-comprehension processes. Lower-level processes function at the surface level of discourse, while higher-level processes function at the inferential level (Grabe, 2009; Kintsch & Rawson, 2005).

Lower-level comprehension processes are those involved in *decoding* and *literal comprehension*. In *decoding*, word recognition takes place either by the direct association of print (*matching*) or print to sound correspondence so that long-term memory representation is accessed (*recoding*). In *literal comprehension*, word meanings are accessed (*lexical access*) and propositions[1] are formed by assembling words together through the use of syntactic rules (*parsing*) (Gagné et al., 1993). Propositional formation depends on bridging inferences and/or anaphora resolution to connect propositions as a coherent whole (Kintsch, 2013; Kintsch & Rawson, 2005). These integrative processes are part of inferential comprehension, that is, "going beyond the idea explicitly stated to integrate, summarize and elaborate on these ideas" (Gagné et al., 1993). However, the meaning of the text is more than the recognition and access of word meanings and the construction of propositions; it entails the provision of a summary of the text and the elaboration of what was read based on background knowledge (Gagné et al., 1993; Kintsch, 2013; Kintsch & Rawson, 2005).

All these processes must be executed within the limited capacity of WM (Daneman & Carpenter, 1980), especially considering reading comprehension as a multilevel architecture that involves both literal and inferential comprehension. In fact, "depending on the reader's interaction with a given level [lower or higher] working memory capacity is differentially involved" (Alptekin & Erçetin, 2009, p. 628), in the sense that literal comprehension demands less working memory resources in relation to inferential comprehension (Just & Carpenter, 1992). Additionally, Just and Carpenter (1992) explain that "when the task demands are high enough to strain capacity, individuals with smaller working memory capacity should be less able to perform computations quickly or store the intermediate products" (p. 143).

With that in mind, one might wonder how readers are able to construct a coherent mental representation considering the limited resources for storage and processing the aforementioned processes involved in reading (Tomitch, 2003). In addition to that, reading a hypertext might overload readers' WM given that readers must decide which links to click and integrate information from all the texts read in the search into the mental representation (DeStefano & LeFevre, 2007). Similarly, WMC in inexperienced digital readers might be overloaded once they may not have fully acquired *navigation, integration and evaluation* skills for hypertext reading (Salmerón et al., 2018). More specifically, 1) readers must know how to select hyperlinks and web pages in the vast array of information online; 2) they must also be able to "integrate multiple pieces of information and multiple presentation formats (texts from different web pages, text and animations)" (p. 91); and 3) they must be able to critically evaluate information found online in terms of quality and reliability (Salmerón et al., 2018). Having said that, this study investigated multitasking in hypertext reading[2], under the speculation that onscreen reading would place more demands on working memory.

Moreover, results from the literature investigating the relationship between WMC and reading times have found longer reading times for lower-span than for higher-span readers, especially in more demanding situations (e.g. King & Just, 1991; Linderholm et al., 2008; Osaka & Osaka,

[1] Propositions are idea units and resemble the meaning of the text as it is constructed by the reader (Kintsch, 2013).
[2] Given the pandemic of COVID-19, in which experiments had to be conducted online, this study was unable to compare paper reading to onscreen reading.

2002). In the present study we expect to find longer reading times for lower-spans in the more demanding condition of reading while listening to lyrical songs.

The Study

Considering that 1) working memory is resource-limited and highly deployed in complex cognitive tasks such as reading (Baddeley, 2012; Just & Carpenter, 1992); 2) reading involves both lower-level (literal comprehension) and higher-level comprehension processes (inferential comprehension) (Gagné et al., 1993; Grabe, 2009); 3) the fact that working memory resources are differently deployed in each level – literal comprehension relies on more automatized processes such as word decoding, lexical access and syntactic parsing, while inferential comprehension relies on integrating text parts, constructing main ideas and bringing background knowledge to elaborate on the text ideas (Alptekin & Erçetin, 2009; Gagné et al., 1993; Just & Carpenter, 1992); in addition to the fact that 4) the phonological component of working memory is highly deployed both in reading and listening to lyrical music (Baddeley, 2012; Vasilev et al., 2018), we hypothesized that literal comprehension is not predicted by working memory capacity nor condition (Hypothesis 1); inferential comprehension is predicted by working memory capacity and condition, and there is an interaction between the two variables (Hypothesis 2); and reading times are predicted by working memory capacity and condition (Hypothesis 3).

Method

Design

This cross-sectional experimental study (Dörnyei, 2007) adopted a between-subject design consisting of two conditions featuring both a primary task (hypertext reading) and a secondary task (music listening) to simulate a multitasking situation. The control condition consisted of reading a hypertext whilst listening to non-lyrical music, while the experimental condition consisted of reading hypertext whilst listening to lyrical music.

Participants

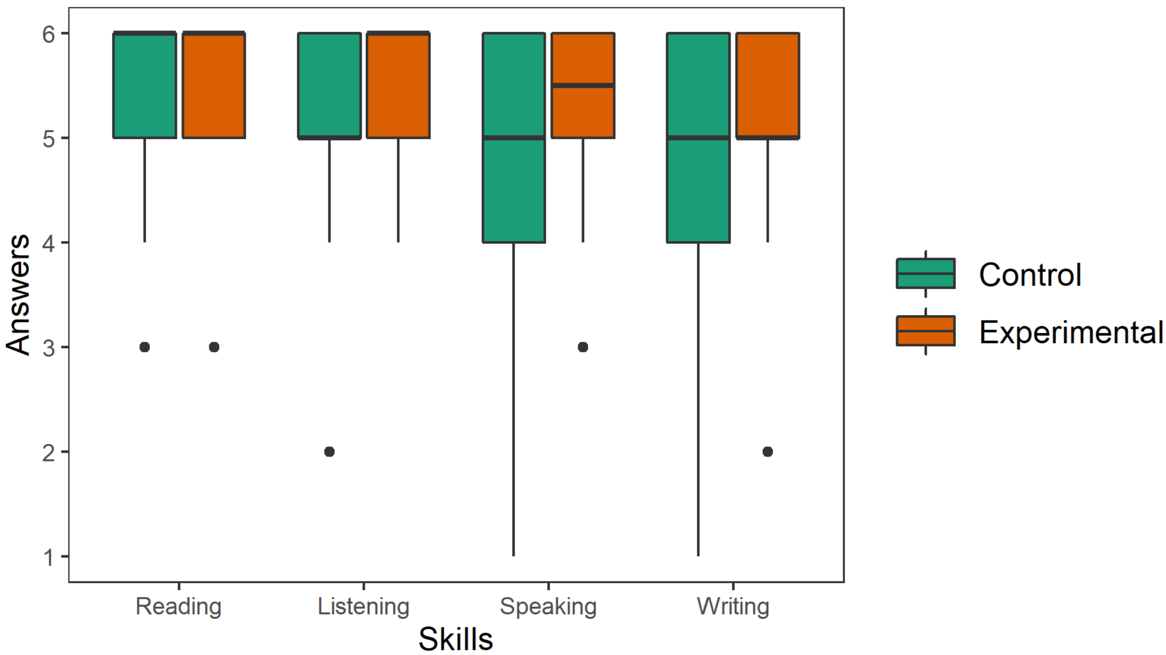
Participants were proficient bilinguals from several regions of Brazil. The most expressive population were students from English undergraduate and graduate programs from *Universidade Federal de Santa Catarina*, and English undergraduate students from *Universidade Federal do Rio Grande do Sul*. Invitations were made through a social network app *(Instagram)* and e-mail. Sample size calculations[3] showed that at least sixty-seven participants were needed for statistical robustness. Sixty-five adults between the ages of 19 and 62 years (M = 28.87; SD = 8.20), 50 females and 15 males, participated in the study. Gender was balanced between the control (24 females, 7 males) and experimental (26 females, 8 males) condition.

[3] Sample size was calculated with the R package pwr (Champely, 2020) considering a medium effect size ($f^2 = .15$), $\alpha = .05$ and power of .8.

Participants' proficiency was self-rated using the *Language Experience and Proficiency Questionnaire* (*QuExPLi*) (Scholl & Finger, 2013), detailed in the *Instruments and Procedures for Data Collection* below. Participants' responses were inspected to explore possible differences in proficiency between the two groups (control and experimental), and the analysis revealed that both groups were balanced in terms of self-rated proficiency (Figure 1), suggesting, thus, that proficiency did not influence the main results of this study.

Figure 1

Participants' Self-rated Proficiency



Note: both groups were balanced in terms of self-rated proficiency

Ethics Review Board

This project was submitted to the Ethics Review Board from *Universidade Federal de Santa Catarina*, in compliance with Resolutions 466/12 and 510/16 of the National Council of Health, from the Brazilian Ministry of Health and was approved under protocol 4.688.798 of May 4, 2021. Consent was given by participants selecting the box stating acceptance of terms for data collection and use. Consent was also considered by the total execution of the tasks contained in this study, following *Ofício Circular nº 2/2021/CONEP/SECNS/MS* of February 14, 2021. Participants received a signed copy of the consent form in their e-mail.

Apparatus

The experimental session was carried out remotely given the pandemic of the COVID-19, using an online platform in development by the second author of this study. The platform, programmed in JavaScript, is called Lapsi – *Laboratório de Psicolinguística na Web* – and will soon be available for researchers in the field of Psycholinguistics, Applied Linguistics and related areas at http://lapsi.davi.solutions/. The platform is a repository of digital instruments of data collection that can be reused and customized. For the present study, all the instruments described in the following sections were implemented in the platform and will be available for reuse in future studies. Data collected in the platform is stored in a database and can be downloaded in JSON or CSV formats for analysis. Participants are required to inform an email and password so that they can have access to their own data.

Piloting

Both a pre-pilot and a pilot study were conducted prior to the actual data collection. Since this study is the first to use the Lapsi platform, it was pre-piloted among fellow researchers in order to verify whether the experiment was running correctly. Having corrected the issues identified at this stage, the pilot study was conducted with a group of five language teachers (age M = 29.4; SD= 7.02) who had attended a teacher training course offered by the first author at *Instituto Federal de Santa Catarina – Câmpus São Lourenço do Oeste*. Data from participants in the pilot study were not included in the final analysis.

Instruments and Procedures for Data Collection

Language Experience and Proficiency Questionnaire

Considering that bilingualism configures a spectrum of experiences (DeLuca et al., 2019; Leivada et al., 2021) that should be taken into account in research with bilinguals (Scholl et al., 2021), participants answered a *Language Experience and Proficiency Questionnaire* (*QuExPLi*) in Brazilian Portuguese (Scholl & Finger, 2013). The adapted questionnaire consists of participants' personal information; their English language background; contributing factors to learning English; and self-rated proficiency. In this article, only the self-rated proficiency is reported. It consisted of four Likert-scale questions asking participants to self-access from 1 to 6 their own English reading, listening, speaking, and writing proficiency. The validity of the *QuExPli* has been recently explored (Scholl et al., 2017, 2021).

Self-Administrable Reading Span Test

The Reading Span Test (Daneman & Carpenter, 1980) consists of reading sixty sentences out loud and trying to recall the last word of each sentence in the order they were originally presented. The sentences are organized into three sets of two, three, four, five, and six sentences, and WMC is thus determined by the level at which the participant fails to recall the last word (Daneman & Carpenter, 1980; Tomitch, 2003).

Most recently, Oliveira et al. (2021) signaled some limitations of the original RST, such as the possibility of retest effects – caused by the repeated use of the same sentences – and the infeasibility of simultaneous application of the test in a single room – given that participants have to read the sentences out loud, which led the researchers to develop a Self-Administrable version of the test. Given these limitations and the fact that this experiment was carried out remotely due to the COVID-19 pandemic, this study adopted Oliveira et al.'s version, which contains 180 sentences, but only 60 are randomly selected every time the experiment is run. The sentences range from 9 to 15 words and are presented in five different levels with three sets in each level, similar to the original test. At the end of each sentence, two buttons are presented containing the end word of the sentence, in which only one word correctly fits the sentence. This procedure is adopted to certify that participants are actually processing the text and not only memorizing the last words of the test and to replace the reading out-loud procedure of the original test. Given the fact that the test was programmed to select words from the 180 sentences, both words given as options – the end words that fit the sentence – might be appropriate for the sentence context (Oliveira et al., 2021). Participants were instructed to select the word they believed was the most appropriate to complete the sentence and recall that word in case that happened. After reading the sentences and selecting the appropriate end-word, participants had to remember the words they had selected and write them in the order they appeared. The test was administered in Brazilian Portuguese to avoid additional language proficiency confounding the results (Tomitch, 2003, 2020) and a familiarization session preceded the actual test.

Text for the Primary Task

The text of this study was controlled in terms of (1) type – expository text; (2) source; (3) number of words; (4) font-size; (5) readability. An expository text on English as a Lingua Franca was extracted from a biannual publication aimed at sharing successful initiatives in language teaching and learning. Only one section of the text was used to control for length. Although the text was displayed in Times New Roman, Font size 12, to facilitate readability (Procailo & Tomitch, 2020), participants' screen size might have varied since it was carried out remotely.

In addition to that, two measures of readability were used in this study: the *Flesch-Kincaid metrics* (Graesser et al., 2004) *and Text Ease and Readability Assessor* (Graesser et al., 2014). After initial analysis showing the original text was somewhat difficult, the text was manipulated, resulting in average in narrativity; average in syntactic simplicity; low word concreteness; average amount of referential cohesion; high in deep cohesion "suggesting more explicit causal relationships as needed by the text. Because of this, it may be easier to comprehend unfamiliar topics" (Coh-Metrix, 2021) (see Graesser et al., 2014 for a full explanation on each component).

The number of hyperlinked information was controlled so that participants had access to a maximum of seven links to avoid distraction (DeStefano & LeFevre, 2007; Procailo & Tomitch, 2020). The hyperlinked information was determined by raters, resulting in seven concepts of the text that were added as a hyperlinked glossary presented as pop-up windows. Participants were told that the hyperlinks contained a detailed account on some of the issues brought by the text.

Auditory Stimuli for the Secondary Task

In both conditions, each participant could choose the song for their reading task, from two options. Two clickable buttons were displayed on the screen, one for each song, in which participants could listen to a 30-second excerpt prior to selecting to which one they would listen whilst reading the hypertext, whose rationale lies on the mixed results in terms of musical preference in comprehension (Johansson et al., 2012; Pereira et al., 2011; Perham & Currie, 2014). This procedure was adopted to simulate the real-life task of selecting a song of their preference for reading. For both conditions, participants were instructed to use an earphone during the experiment and set the volume to feel comfortable during reading.

For the control condition, participants could choose between two songs, retrieved from a playlist for studying – *Binaural Beats: Focus* – available on Spotify and containing 244.568

followers[4]. Binaural Beats are considered to be less distracting than other instrumental pieces that feature acoustic variations that result in distraction (Vasilev et al., 2018). The control group participants chose either ‘Alpha Thoughts 107 Hz – 114 Hz’ (101 *bpms*) or ‘Binaural Alpha Sinus’ (120 *bpms*).

Tempo was manipulated using the software *MixMeister BPM Analyzer*, and ranged from 100 to 120 bpm, which is considered *moderato* in terms of tempo (Fernández-Sotos et al., 2016), under the rationale that slow and soft music does not impair comprehension as fast and loud music does (Thompson et al., 2012).

Participants could also choose between two songs in the experimental condition. Songs were extracted from the playlist containing the most-streamed pop songs nowadays - *Today's Top Hits* – on Spotify (27.433.179 followers). Participants either chose Dua Lipa's ‘Levitating’ (103 *bpms*) or The Weeknd's ‘Save your tears’ (117 *bpms*).

Comprehension Questions

The comprehension questions elaborated for this study followed Pearson and Johnson's (1978) Taxonomy, who classified comprehension questions into *Textually Explicit*, whose answers can be easily identified in the text, and *Textually Implicit*, which involves questions whose answers should not be obviously identified, so that readers may engage in some inferential process. Participants answered five open-ended comprehension questions on the text, based on the taxonomy explained above. The questions were elaborated and rated by the authors of this study resulting in three *Textually Explici*t and two *Textually Implicit* questions. Participants' answers were scored by independent raters and is further detailed in the *Procedures for Data Analysis* below.

In this study, textually explicit questions assessed literal comprehension and textually implicit questions assessed inferential comprehension. Participants could re-read the text while answering the comprehension questions.

Procedures for Data Analysis

The Self-Administrable RST was scored both strictly and leniently given the possible differences among highly correlated scoring methods (Roscioli & Tomitch, 2022). The strict method considers the reading span from the ability to recall at least two out of three sets (Daneman & Carpenter, 1980; Tomitch, 2003). In previous versions of the test, failing to recall three sets would result in the test being terminated (Daneman & Carpenter, 1980; Tomitch, 2003). However, given that this study used the Self-Applicable Reading Span Test (Oliveira et al., 2021), participants proceeded until the end of the test, and their reading span was considered the point at which they were able to recall at least two out of three sets. Similar to Tomitch (2003), the point at which the participant was able to recall at least two sets is taken as the reading span. Additionally, half a point was given to participants who reached one set at the next level. For instance, if a participant scored two sets at the three-sentence level and one set at the four-sentence level, then his/her working memory span would be 3,5. While previous studies considered the exact words in the exact order recalled as a means to score in the RST (Daneman

[4] The number of playlists' followers date from March 27, 2021.

& Carpenter, 1980; Tomitch, 2003), this study disregarded mistakes in word gender (masculine/feminine) and number (singular/plural) since raw data showed that a great number of participants recalled the correct words in the correct order but failed in number and gender markings. The rationale for scoring the RST strictly relies on the assumption that attentional resources are tapped by having participants simultaneously process incoming information at the same time they recall the last words of previously displayed sentences (Daneman & Carpenter, 1980).

The lenient scoring method, on the other hand, considers the total number of words recalled as the reading span (Friedman & Miyake, 2005). While in the strict scoring method participants must recall the words in the exact order they appeared, the lenient scoring method considers the total number of words irrespective of their order in the set. The rationale for scoring the RST leniently relies on the claim that more continuous scoring methods such as the lenient method present higher reliability[5] and more normal distribution[6] of data in relation to strict methods (Friedman & Miyake, 2005). The scores from both methods were used in the analysis of the internal consistency reliability of the Self-Administrable Reading Span Test and the value for Cronbach's Alpha was α = .85, considered good internal consistency.

The comprehension questions were corrected by three independent raters who assigned 1 point per correct answer and 0,5 if the answer was partially correct (Roscioli & Tomitch, 2022). The mean scores assigned by the raters were calculated and used for the statistical analysis. Internal consistency was analyzed with the scores from literal comprehension and inferential comprehension and resulted in α = .72 for Cronbach's Alpha, considered acceptable internal consistency. Last, the *QuExPli* was inspected to investigate possible differences in proficiency between the two groups.

Reading times were measured as the duration that the text for the Primary Task remained on the screen, in other words, as the time elapsed between the click in the button to go to the page the text was presented and the click in the button that showed the next page, in which the subsequent task was presented.

Hypothesis testing was conducted by means of multiple linear regression analyses with the statistical tool R (R Core Team, 2021). Two models were fitted for each response variable, namely literal comprehension, measured by scores from the textually explicit comprehension questions, inferential comprehension, measured by scores from the textually implicit comprehension questions, and reading times, measured as the total time the text for the Primary Task remained visible on the screen. For the six models, the predictor variables were WMC scores and Condition (Control and Experimental), however, for each response variable one model was fitted with the lenient WMC scores and the other one with the strict scores.

---

[5] Reliability refers to the consistency of a research instrument (i.e., a test), in the sense that the same score would be obtained if a person retook the test (Larson-Hall, 2016)

[6] Normal distribution refers to the probability function of how a variable is distributed. In normally distributed data, most observations cluster around the mean, with lower and higher values decreasing symmetrically in probability.

Results

This study investigated whether working memory capacity predicts literal comprehension, inferential comprehension and reading times of a hypertext in proficient bilinguals in a multitasking setting in two conditions – listening to lyrical music (experimental) and listening to non-lyrical music (control). The hypotheses are that 1) literal comprehension is neither predicted by working memory capacity nor condition, that 2) inferential comprehension is predicted by working memory capacity and condition, and there is an interaction between the two variables, and that 3) reading times are predicted by working memory capacity and condition.

Literal Comprehension

We hypothesized that literal comprehension would not be predicted by WMC and Condition – listening to lyrical music during reading in the experimental condition and listening to non-lyrical music while reading in the control condition. No statistically significant effects were observed for WMC, condition, nor interaction, irrespective of the scoring method used for the WMC task, supporting hypothesis 1 (Table 1). In other words, the linear models showed that neither WMC (scored leniently or strictly) nor condition affected literal comprehension of a hypertext (Table 1).

Table 1

Estimates of WMC (A) Lenient Method and (B) Strict Method and Condition as Predictors of Literal Comprehension

(a)

Names	*β*	*p*	95% CI
Intercept	0.48	0.02*	[0.084, 0.87]
WMC (Lenient)	0.01	0.38	[-0.0047, 0.012]
Condition	0.13	0.62	[-0.39, 0.65]
Interaction	-0.01	0.66	[-0.014, 0.009]

(b)

Names	*β*	*p*	95% CI
Intercept	0.60	< 0.01***	[0.37, 0.83]
WMC (Strict)	0.02	0.64	[-0.045, 0.074]
Condition	0.11	0.48	[-0.19, 0.4]
Interaction	-0.03	0.48	[-0.11, 0.051]

Note. No statistically significant effect nor interaction between predictor variables. R^2 were (a) 0.15 and (b) 0.009. CI = Confidence Interval for *β*.

* $p < .05$. *** $p < .001$

Inferential Comprehension

We also hypothesized that WMC and Condition – listening to non-lyrical music (control), as compared to lyrical music (experimental) – would predict inferential comprehension of a hypertext. With the lenient scoring method, a single statistically significant effect of WMC was observed (Table 2a). However, with the strict scoring method, both predictors (WMC, Condition) and their interaction showed statistically significant effects (Table 2b). Thus, Hypothesis 2 is supported by the data only if considering the strict method. It is worth noticing, however, that the R^2 of the model with the strict scoring method is lower than the R^2 of the model with the lenient scoring method ($R^2 = .116$; $R^2 = .123$, respectively), which means that the lenient scoring method results in a linear model that is slightly better at explaining the variability in the data.

Table 2

Estimates of WMC (A) Lenient Method and (B) Strict Method and Condition as Predictors of Inferential Comprehension

(a)

Names	β	p	95% CI
Intercept	0.05	0.81	[-0.34, 0.44]
WMC	0.01	0.01*	[0.01, 0.03]
Condition	0.45	0.09	[-0.07, 0.97]
Interaction	-0.01	0.12	[-0.03, 0.03]

(b)

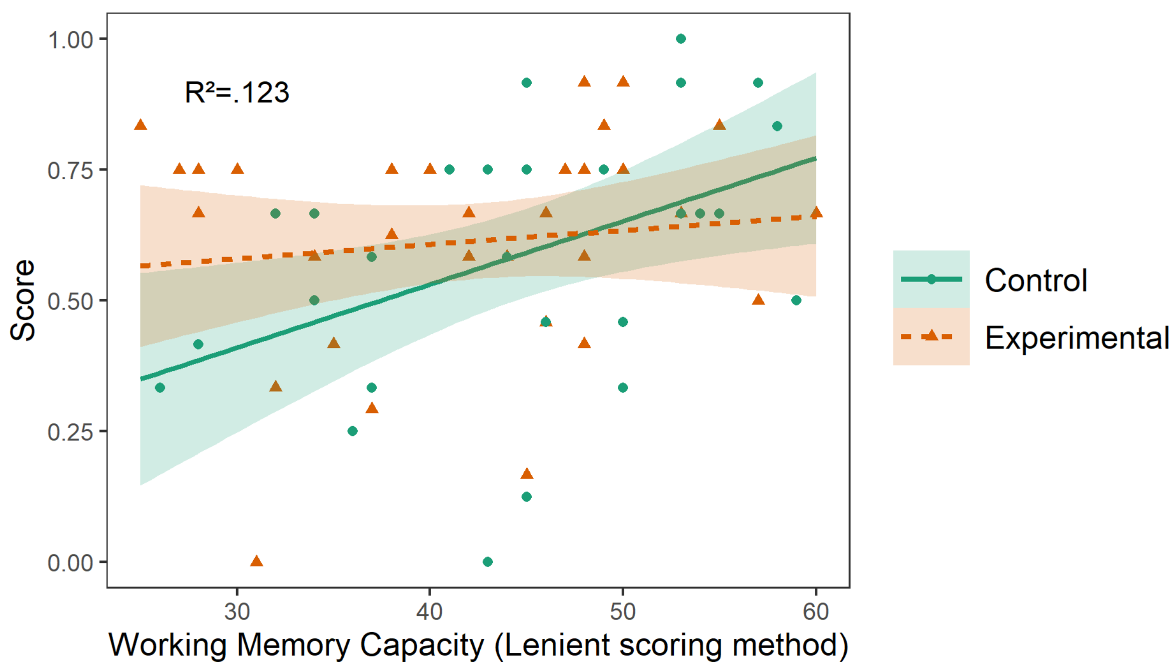
Names	β	p	95% CI
Intercept	0.29	0.02*	[0.06, 0.52]
WMC	0.08	0.01*	[0.03, 0.14]
Condition	0.37	0.02*	[0.08, 0.67]
Interaction	-0.09	0.02*	[-0.18, -0.02]

Note. R^2 were (a) 0.12 and (b) 0.12. CI = Confidence Interval for β

$^* p < .05$

Figure 2

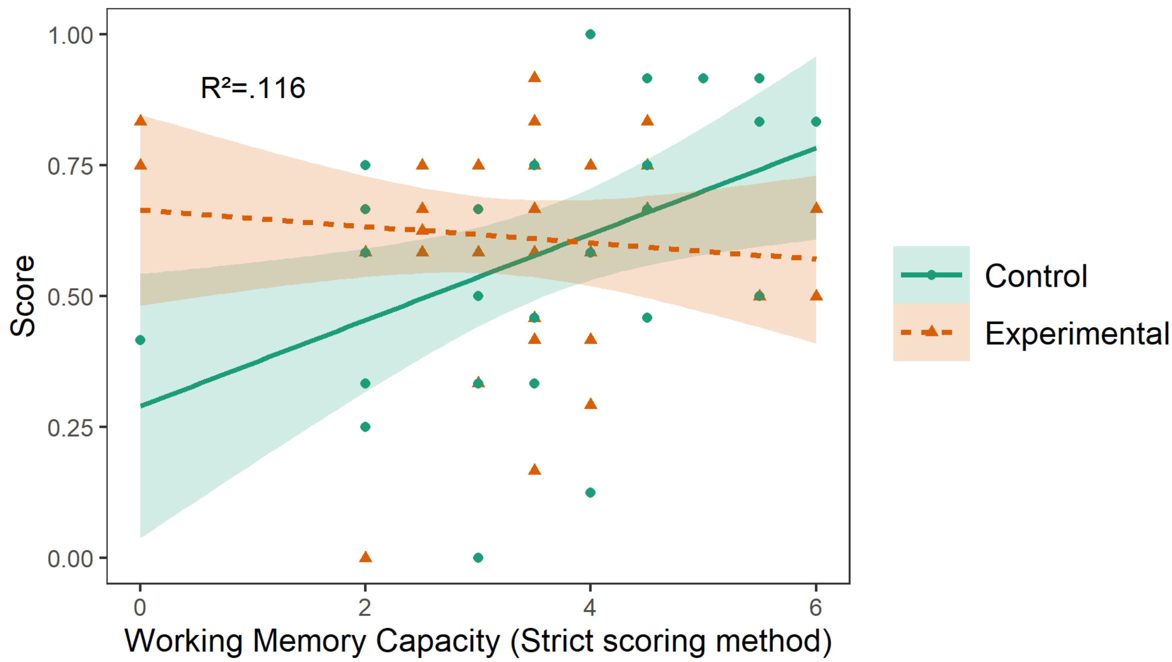
Inferential Comprehension Predicted by WMC (Lenient Method) and Condition



Note. Regression slopes according to Condition.

Figure 3

Inferential Comprehension Predicted by WMC (Strict Method) and Condition



Note. Regression slopes according to Condition.

Reading Times

Finally, we hypothesized that reading times would be predicted by WMC and Condition. No statistically significant effects were observed for WMC, condition, nor interaction, irrespective of the scoring method used for the WMC task. Thus, the data do not support hypothesis 3 (Table 3). In other words, the linear models showed that neither WMC (scored leniently or strictly) nor condition affected the time participants spent reading a hypertext.

Table 3

Estimates of WMC (a) Lenient Method and (b) Strict Method and Condition as Predictors of Reading Times

(a)

Names	*β*	*p*	95% CI
Intercept	6.07	$< 0.01^{***}$	[2.98, 9.17]
WMC (Lenient)	-0.07	0.12	[-0.15, 0.02]
Condition	-2.98	0.18	[-7.39, 1.42]
Interaction	0.09	0.14	[-0.03, 0.22]

(b)

Names	*β*	*p*	95% CI
Intercept	4.50	$< 0.01^{***}$	[2.65, 6.35]
WMC (Strict)	-0.29	0.34	[-0.91, 0.32]
Condition	0.17	0.90	[-2.57, 2.91]
Interaction	0.00	0.99	[-0.95, 0.95]

Note. No statistically significant effect nor interaction between predictor variables. R^2 were (a) 0.046 and (b) 0.027. CI = Confidence Interval for *β*. *** p < .001

Discussion

Overall, as mentioned, reading times were not affected by working memory capacity nor condition. Our higher-span readers spent as much time reading the texts as lower-span ones in both conditions: while concomitantly listening to lyrical songs and while listening to non-lyrical music. As advocated by Just and Carpenter (1992) differences between the performances of high- and low-span individuals are seen when the task is demanding enough so as to trigger WM resources, leading to a trade-off between storage and processing when capacity limits are exceeded. Our results show that listening to lyrical songs while reading the text did not take our participants (both higher- and lower-spans) to spend more time on the text than while reading and listening to non-lyrical music. These findings are corroborated by Schurer et al. (2020) who found no differences in reading times between coherent and incoherent texts in an investigation of the impact of WMC and prior knowledge on readers' attention and comprehension.

The statistically insignificant effects obtained from the linear models for WMC, condition and interaction can be interpreted as evidence that literal comprehension demands little resources from readers' working memory. The lack of effect of WMC on literal comprehension lends support to the *total capacity explanation* of individual differences which assumes that "capacity limitations affect performance only if the resource demands of the task exceed the available supply" (Just & Carpenter, 1992, p. 145). In other words, literal comprehension might not have deployed all available working memory resources, especially considering participants' higher levels of proficiency (evidenced in the exploratory analysis), which is intimately related to grammatical knowledge used in syntactic parsing and vocabulary knowledge needed for lexical access (Gagné et al., 1993; Grabe, 2009). Additionally, these results seem to be in line with Alptekin and Erçetin's (2010) findings that literal comprehension heavily relied on language proficiency, decoding, and syntactic parsing.

These findings might suggest that lower-level processes and higher-level processes function separately, as evidenced by previous studies (Alptekin & Erçetin, 2011; Hannon, 2012). In detail, Alptekin and Erçetin (2011) proposed that "higher-order and lower-order reading operations reflect independent cognitive systems at work" (p. 257) based on their findings that domain knowledge did not contribute to literal comprehension. Similarly, Hannon (2012) assumed that lower-level and higher-level processes function independently in adult readers, based on evidence from a structural equation model that tested the potential relationship among lower- and higher-level processes and working memory as sources of individual differences in reading comprehension altogether (Hannon, 2012). However, the assumption that lower-level and higher-level processes work independently is both speculative (e.g. Alptekin & Erçetin, 2011) and preliminary (e.g. Hannon, 2012), thus it deserves further investigation.

It is important to note that the results of this study find support both in the domain-general and the domain-specific views of working memory (Logie et al., 2021; Wen, 2015; Wen & Li, 2019), following the emerging trend to integrate several models of working memory (Logie et al., 2021; Wen & Schwieter, 2022). More specifically, the results can be explained by the Working Memory and Executive Attention perspective (Engle, 2018; Mashburn et al., 2021); and the Multicomponent Model (Baddeley, 2017).

The significant effects obtained from inferential reading whilst listening to lyrical music were only found in the strict scoring method and converge with the claim that "the ability to do a cognitively demanding task […] while trying to recall the last word of each sentence is believed to require attention, and therefore the strict scoring method would show the limitations of WMC" (Roscioli, 2017, p. 90). This claim seems to be in line with Engle and colleagues' view of working memory as the system responsible for maintaining goal-relevant information in an active state, which has been evidenced by complex span tasks, such as the RST (Engle, 2018; Mashburn et al., 2021). In fact, WMC relies on domain-general executive attention to keep goal-relevant information in an active state preventing distractions. Thus, our results demonstrate that WMC, which is dependent on attentional control, may be compromised during multitasking.

Nevertheless, it seems the Multicomponent Model of Working Memory also explains the results obtained, given its modular nature (Baddeley, 2017). The domain-specific component known as the phonological loop, which is held accountable for storing and manipulating verbal information (Baddeley, 2012; Linck et al., 2014) is believed to be highly deployed while reading and listening to music, given that "speech sounds automatically gain access to phonological loop" (Vasilev et al., 2018, p. 571), as evidenced by a series of experiments that explored the effects of background music on phonological short-term memory by having participants recall a series of visually presented verbal items whilst listening to either instrumental or vocal (lyrical) music. Even though reading is more complex than recalling visually presented items, it somehow resembles the process of keeping the order of words and their syntactic relations for constructing the mental representation (Vasilev et al., 2018). To be more specific, the surface level of comprehension involves word and parsing processes in order to form idea units or propositions (Kintsch, 1998, 2013; Kintsch & Rawson, 2005), thus, "forming these units must also involve establishing and keeping track of the order of words in the sentence, as well as their syntactic relationships" (Vasilev et al., 2018, p. 571). With that in mind, in multitasking the phonological component of working memory (the loop) may be overloaded by both the concurrent streams of information coming from the input of the primary task (reading), and vocal and acoustical information from the secondary task.

All in all, these results seem to provide evidence for the claim that working memory is both domain-general, in terms of executive control functioning, and domain-specific, concerning phonological input (see Wen & Schwieter, 2022 for a full account). These authors suggest, therefore, that "it may be best to conceive working memory as a multicomponent system that consists of both domain-specific storage buffers and domain-general executive control functions […]. Therefore, neither a completely domain-general nor a completely domain-specific view of working memory holds" (Wen & Schwieter, 2022, p. 913).

Conclusions

The study reported in this article aimed at investigating whether working memory capacity predicted literal and inferential comprehension of proficient bilinguals in a multitasking setting, elicited by the use of songs during a reading task – lyrical pop songs for the experimental group and non-lyrical songs for the control group. Results of multiple linear regression showed that working memory significantly predicts inferential, but not literal comprehension, and only when participants were listening to lyrical music. Thus, these results suggest that the construction of a mental representation in hypertext reading might be compromised in a multitasking setting, since inferences may not be drawn due to lack of working memory resources to support them.

These results seem to be in consonance with the claim that multitasking is in fact distraction (Aagaard, 2019), at least for higher levels of comprehension. The author challenges the views adopted by the techno-optimists who advocate that there is a new generation able to perform several tasks at the same time and claims that "multitasking is not a matter of attention divided, but of attention diverted" (p. 88). Therefore, our findings have clear pedagogical implications for reading instruction, in the sense that students should be encouraged to opt for non-lyrical songs while reading expository texts in the digital format.

This study has some limitations that could be addressed by future studies. The first regards the lack of a reading on paper condition so that it could be compared to onscreen reading. Such limitation is due to the pandemic of COVID-19 in which the study had to be carried out respecting social distancing. The second concerns the lack of a reading-in-silence group to compare with reading while listening to non-lyrical music and reading while listening to lyrical music. Given "the omnipresence of sound within the society we live in [and the] (...) rare occurrence to undertake a cognitive task in a quiet environment" (Elliott et al., 2022, p. 307), we decided not to include a reading-in-silence group, since our goal was to conduct a study which simulated a situation in which individuals might experience in their everyday lives. The third limitation regards the conceptualization of inferential comprehension. We considered inferential comprehension questions those rated as *textually implicit* questions (Pearson & Johnson, 1978). Previous studies, nevertheless, had a more thorough account of inferences (to mention a few Procailo & Tomitch, 2020; Roscioli & Tomitch, 2022). For instance, Procailo and Tomitch (2020) had two independent raters categorize inferences "associations, evaluative comments, elaborative inferences, predictive inferences, reinstatement inferences, metacognitive comments, paraphrases, […] text repetitions, […] summarizations (main idea of the paragraph) and misunderstandings" (Procailo & Tomitch, 2020, pp. 331-332) (see Linderholm & van den Broek, 2002 for a full account). This limitation could be addressed by future studies investigating inferential comprehension in multitasking.

Another limitation is the use of the RST, a verbal task, as a single measure of WMC. The use of the RST is based on the assumption that despite being a verbal task the test is sensitive to individual differences in domain-general executive functions of WM. However, even with this assumption, it is also assumed that individual differences in verbal proficiency and in the domain-specific buffers of WM also influence the test. Thus, using tests that involve other [domains], such as the Operation Span or the Symmetry Span (Kane et al., 2004) can better account for influences from domain-specific factors on the task.

References

Aagaard, J. (2019). Multitasking as distraction: A conceptual analysis of media multitasking research. *Theory & Psychology*, *29*(1), 87-99. https://doi.org/10.1177/0959354318815766

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*(4), 627–639. https://doi.org/10.1016/j.system.2009.09.007

Alptekin, C., & Erçetin, G. (2010). The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading. *Journal of Research in Reading*, *33*(2), 206-219. https://doi.org/10.1111/j.1467-9817.2009.01412.x

Alptekin, C., & Erçetin, G. (2011). Effects of Working Memory Capacity and Content Familiarity on Literal and Inferential Comprehension in L2 Reading. *TESOL Quarterly*, *45*(2), 235-266. https://doi.org/10.5054/tq.2011.247705

Backer, K. C., & Bortfeld, H. (2021). Characterizing Bilingual Effects on Cognition: The Search for Meaningful Individual Differences. *Brain Sciences*, *11*(1), 81. https://doi.org/10.3390/brainsci11010081

Baddeley, A. D. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, *63*(1), 1-29. https://doi.org/10.1146/annurev-psych-120710-100422

Baddeley, A. D. (2017). Modularity, working memory and language acquisition. *Second Language Research*, *33*(3), 299-311. https://doi.org/10.1177/0267658317709852

Baddeley, A. D., Hitch, G., & Allen, R. (2021). A Multicomponent Model of Working Memory. In R. H. Logie, V. Camos, & N. Cowan (Eds.), *Working Memory: State of the Science* (pp. 10-43). Oxford University Press. https://doi.org/10.1093/oso/9780198842286.003.0002

Bhatia, E. T. K., Ritchie, W. C., Horn, E. L. R., Ward, G., Kouwenberg, E. S., Singler, J. V., & Doughty, C. J. (2012). *The Handbook of Bilingualism and Multilingualism* (2nd ed.). Blackwell Publishing.

Bowman, L. L., Levine, L. E., Waite, B. M., & Gendron, M. (2010). Can students really multitask? An experimental study of instant messaging while reading. *Computers & Education*, *54*(4), 927-931. https://doi.org/10.1016/j.compedu.2009.09.024

Bråten, I., Braasch, J. L. G., & Salmerón, L. (2020). Reading Multiple and Non-Traditional Texts. In E. B. Moje, P. P. Afflerbach, P. Enciso, & N. K. Lesaux, *Handbook of Reading Research, Volume V* (1st ed., pp. 79-98). Routledge. https://doi.org/10.4324/9781315676302-5

Champely, S. (2020). *pwr: Basic Functions for Power Analysis* (R package version 1.3-0) [R]. https://CRAN.R-project.org/package=pwr

Cho, K. W., Altarriba, J., & Popiel, M. (2015). Mental Juggling: When Does Multitasking Impair Reading Comprehension? *The Journal of General Psychology*, *142*(2), 90-105. https://doi.org/10.1080/00221309.2014.1003029

Clinton-Lisell, V. (2021). Stop multitasking and just read: Meta-analyses of multitasking's effects on reading performance and reading time. *Journal of Research in Reading*, *44*(4), 787-816. https://doi.org/10.1111/1467-9817.12372

Coh-Metrix. (2021). COH-METRIX. http://cohmetrix.com/

Cowan, N., Morey, C. C., & Naveh-Benjamin, M. (2021). An embedded-processes approach to working memory: How is it distinct from other approaches, and to what ends? In R. H. Logie, V. Camos, & N. Cowan (Eds.) *Working memory: State of the science.* (pp. 44–84). Oxford University Press.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. Psychonomic Bulletin & Review, 12(5), 769–786. https://doi.org/10.3758/BF03196772

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450-466. https://doi.org/10.1016/S0022-5371(80)90312-6

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422-433. https://doi.org/10.3758/BF03214546

DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences*, *116*(15), 7565-7574. https://doi.org/10.1073/pnas.1811513116

DeStefano, D., & LeFevre, J. A. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior*, *23*(3), 1616–1641. https://doi.org/10.1016/j.chb.2005.08.012

Dörnyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford University Press.

Elliott, E. M., Bell, R., Gorin, S., Robinson, N., & Marsh, J. E. (2022). Auditory distraction can be studied online! A direct comparison between in-Person and online experimentation. *Journal of Cognitive Psychology*, *34*(3), 307–324. https://doi.org/10.1080/20445911.2021.2021924

Engle, R. W. (2018). Working Memory and Executive Attention: A Revisit. *Perspectives on Psychological Science*, *13*(2), 190-193. https://doi.org/10.1177/1745691617720478

Fernández-Sotos, A., Fernández-Caballero, A., & Latorre, J. M. (2016). Influence of Tempo and Rhythmic Unit in Musical Emotion Regulation. *Frontiers in Computational Neuroscience*, *10*. https://doi.org/10.3389/fncom.2016.00080

Friedman, N. P., & Miyake, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods*, *37*(4), 581-590. https://doi.org/10.3758/BF03192728

Gagné, E., Yekovich, C. W., & Yekovich, F. (1993). Reading. In *Cognitive Psychology of School Learning* (2nd ed., p. 512). Allyn & Bacon.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practive*. Cambridge University Press.

Graesser, A. C., McNamara, D. S., Cai, Z., Conley, M., Li, H., & Pennebaker, J. (2014). Coh-Metrix Measures Text Characteristics at Multiple Levels of Language and Discourse. *The Elementary School Journal*, *115*(2), 210-229. https://doi.org/10.1086/678293

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. https://doi.org/10.3758/BF03195564

Hannon, B. (2012). Understanding the Relative Contributions of Lower-Level Word Processes, Higher-Level Processes, and Working Memory to Reading Comprehension Performance in Proficient Adult Readers. *Reading Research Quarterly*, *47*(2), 125–152. https://doi.org/10.1002/RRQ.013

In'nami, Y., Hijikata, Y., & Koizumi, R. (2021). Working Memory Capacity and L2 Reading: A Meta-Analysis. *Studies in Second Language Acquisition*, 1-26. https://doi.org/10.1017/S0272263121000267

Johansson, R., Holmqvist, K., Mossberg, F., & Lindgren, M. (2012). Eye movements and reading comprehension while listening to preferred and non-preferred study music. *Psychology of Music*, *40*(3), 339-356. https://doi.org/10.1177/0305735610387777

Just, M. A., & Carpenter, P. A. (1992). *A capacity theory of comprehension: Individual differences in working memory*. 29. https://doi.org/10.1037/0033-295X.99.1.122

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189-217. https://doi.org/10.1037/0096-3445.133.2.189

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*(5), 580-602. https://doi.org/10.1016/0749-596X(91)90027-H

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Kintsch, W. (2013). Revisiting the Construction–Integration Model of Text Comprehension and Its Implications for Instruction. In N. Unrau (Ed.), *Theoretical Models and Processes of Reading* (6th ed., pp. 807-839). International Reading Association. https://doi.org/10.1598/0710.32

Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 211-226). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470757642.ch12

Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, *9*(2), 119-135. https://doi.org/10.1017/S1366728906002483

Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R* (Second Edition). Routledge.

Leeser, M. J., & Herman, E. (2022). Methodological Issues in Research on Working Memory and L2 Reading Comprehension. In J. W. Schwieter & Z. Wen (Eds.), *The Cambridge Handbook of Working Memory and Language* (1st ed., pp. 613-634). Cambridge University Press.

Leivada, E., Westergaard, M., Duñabeitia, J. A., & Rothman, J. (2021). On the phantom-like appearance of bilingualism effects on neurocognition: (How) should we proceed? *Bilingualism: Language and Cognition*, *24*(1), 197-210. https://doi.org/10.1017/S1366728920000358

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, *21*(4), 861-883. https://doi.org/10.3758/s13423-013-0565-2

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. https://doi.org/10.1037/0022-0663.94.4.778

Linderholm, T., Cong, X., & Zhao, Q. (2008). Differences in Low and High Working-Memory Capacity Readers' Cognitive and Metacognitive Processing Patterns as a Function of Reading for Different Purposes. *Reading Psychology*, *29*(1), 61-85. https://doi.org/10.1080/02702710701568587

Logie, R. H., Belletier, C., & Doherty, J. M. (2021). Integrating Theories of Working Memory. In R. H. Logie, V. Camos, & N. Cowan (Eds.), *Working Memory: State of the science* (pp. 389-430). Oxford University Press. https://doi.org/10.1093/oso/9780198842286.003.0014

Logie, R. H., Camos, V., & Cowan, N. (Eds.). (2021). *Working memory: State of the science*. Oxford University Press.

Mashburn, C. A., Tsukahara, J. S., & Engle, R. W. (2021). Individual Differences in Attention Control: Implications for the Relationship Between Working Memory Capacity and Fluid Intelligence. In R. H. Logie, V. Camos, & N. Cowan (Eds.), *Working Memory: The state of the science* (pp. 175–211). Oxford University Press. https://doi.org/10.1093/oso/9780198842286.003.0007

May, K. E., & Elder, A. D. (2018). Efficient, helpful, or distracting? A literature review of media multitasking in relation to academic performance. *International Journal of Educational Technology in Higher Education*, *15*(1), 13. https://doi.org/10.1186/s41239-018-0096-z

Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current Directions in Psychological Science*, *21*, 8-14. https://doi.org/10.1177/0963721411429458

Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science*, *10*(6), 719-726. https://doi.org/10.1111/j.1467-7687.2007.00623.x

Oliveira, D. A., Woelfer, S., & Tomitch, L. (2021). Construindo um Teste de Capacidade de Leitura Autoaplicável com apresentação aleatória de estímulos [Building a Self-Administrable Reading Span Test with Random Stimuli Presentation]. *Revista Neuropsicologia Latinoamericana*, *13*(1), 62-76. https://doi.org/10.5579/rnl.2016.0560

Osaka, N., & Osaka, M. (2002). Individual Differences in Working Memory during Reading with and without Parafoveal Information: A Moving-Window Study. *The American Journal of Psychology*, *115*(4), 501. https://doi.org/10.2307/1423525

Paap, K. (2019). The Bilingual Advantage Debate: Quantity and Quality of the Evidence. In J. W. Schwieter & M. Paradis (Eds.), *The Handbook of the Neuroscience of Multilingualism* (1st ed., pp. 701–735). Wiley. https://doi.org/10.1002/9781119387725.ch34

Pashler, H., Kang, S. H. K., & Ip, R. Y. (2013). Does Multitasking Impair Studying? Depends on Timing: Does multitasking impair studying? *Applied Cognitive Psychology*, *27*(5), 593-599. https://doi.org/10.1002/acp.2919

Pearson, P. D., & Johnson, D. D. (1978). *Teaching Reading Comprehension*. Holt, Rinehart & Winston.

Pereira, C. S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L., & Brattico, E. (2011). Music and Emotions in the Brain: Familiarity Matters. *PLoS ONE*, *6*(11), e27241. https://doi.org/10.1371/journal.pone.0027241

Perham, N., & Currie, H. (2014). Does listening to preferred music improve reading comprehension performance?: Preferred music and reading comprehension. *Applied Cognitive Psychology*, *28*(2), 279–284. https://doi.org/10.1002/acp.2994

Poarch, G. J., & Bialystok, E. (2015). Bilingualism as a model for multitasking. *Developmental Review*, *35*, 113–124. https://doi.org/10.1016/j.dr.2014.12.003

Poarch, G. J., & Krott, A. (2019). A Bilingual Advantage? An Appeal for a Change in Perspective and Recommendations for Future Research. *Behavioral Sciences*, *9*(9), 95. https://doi.org/10.3390/bs9090095

Pollard, M. A., & Courage, M. L. (2017). Working memory capacity predicts effective multitasking. *Computers in Human Behavior*, *76*, 450-462. https://doi.org/10.1016/j.chb.2017.08.008

Procailo, L., & Tomitch, L. M. B. (2020). Investigating Digital Reading in L2 to Criticize and to Summarize: Working Memory Capacity and Reading Purpose Influencing Strategy. *Lingüística y Literatura*, *41*(78), 320–351. https://doi.org/10.17533/udea.lyl.n78a13

R Core Team. (2021). *R: A language and environment for statistical computing* [R]. R Foundation for Statistical Computing. https://www.R-project.org/.

Roscioli, D. C. (2017). *The Relationship Between Technical High School Brazilian Students' Working Memory Capacity, Pre-Reading Activities, and Inference Generation in Reading Comprehension in L2* [PhD Dissertation, Universidade Federal de Santa Catarina]. https://repositorio.ufsc.br/xmlui/handle/123456789/183399

Roscioli, D. C., & Tomitch, L. M. B. (2022). The Influence of Working Memory Capacity on Inference Generation and Reading Comprehension. *Alfa: Revista de Linguística (São José Do Rio Preto)*, *66*, e13543. https://doi.org/10.1590/1981-5794-e13543

Salmerón, L., Strømsø, H. I., Kammerer, Y., Stadtler, M., & van, P. (2018). Comprehension processes in digital reading. In M. Barzillai, J. Thomson, S. Schroeder, & P. van den Broek (Eds.), *Learning to read in a digital world* (pp. 91-120). John Benjamins Publishing Company. https://doi.org/10.1075/swll.17

Salvucci, D. D., & Taatgen, N. A. (2011). *The multitasking mind*. Oxford University Press.

Scholl, A. P., & Finger, I. (2013). Elaboração de um questionário de histórico da linguagem para pesquisas com bilíngues [Development of a Language Background Questionnaire for Research with Bilinguals]. *Nonada: Letras Em Revista*, *2*(21), 1-17.

Scholl, A. P., Finger, I., & Fontes, A. B. A. da L. (2017). Fatores de experiência linguística associados à proficiência autoavaliada por usuários de inglês como língua adicional [Language Experience Factors Associated to Self-Reported Proficiency by Users of English as an Additional Language]. *Letrônica*, *10*(2), 689–699. https://doi.org/10.15448/1984-4301.2017.2.26180

Scholl, A. P., Fontes, A. B. A. da L., & Finger, I. (2021). Can bilinguals rate their proficiency accurately in a language background questionnaire? A correlation between self-rated and objective proficiency measures. *Revista Da Anpoll*, *52*(1), 142-161. https://doi.org/10.18309/ranpoll.v52i1.1506

Schurer, T., Opitz, B., & Schubert, T. (2020). Working Memory Capacity but Not Prior Knowledge Impact on Readers' Attention and Text Comprehension. *Frontiers in Education*, *5*, 26. https://doi.org/10.3389/feduc.2020.00026

Shin, J. (2020). A meta-analysis of the relationship between working memory and second language reading comprehension: Does task type matter? *Applied Psycholinguistics*, *41*(4), 873–900. https://doi.org/10.1017/S0142716420000272

Sörman, D. E., Josefsson, M., Marsh, J. E., Hansson, P., & Ljungberg, J. K. (2017). Longitudinal effects of bilingualism on dual-tasking. *PLOS ONE*, *12*(12), e0189299. https://doi.org/10.1371/journal.pone.0189299

Sparks, R. L. (2019). Why reading is a challenge for U.S. L2 learners: The impact of cognitive, ecological, and psychological factors in L2 comprehension. *Foreign Language Annals*, *52*(4), 727-743. https://doi.org/10.1111/flan.12432

Sparks, R. L. (2021). Identification and Characteristics of Strong, Average, and Weak Foreign Language Readers: The Simple View of Reading Model. *The Modern Language Journal*, *105*(2), 507-525. https://doi.org/10.1111/modl.12711

Sparks, R. L., Patton, J., & Luebbers, J. (2019). Individual differences in L2 achievement mirror individual differences in L1 skills and L2 aptitude: Crosslinguistic transfer of L1 to L2 skills. *Foreign Language Annals*, *52*(2), 255-283. https://doi.org/10.1111/flan.12390

Thompson, W. F., Schellenberg, E. G., & Letnic, A. K. (2012). Fast and loud background music disrupts reading comprehension. *Psychology of Music*, *40*(6), 700-708. https://doi.org/10.1177/0305735611400173

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

Tomitch, L. M. B. (2020). Aspectos importantes para a construção, aplicação e interpretação dos resultados de um instrumento de medida da capacidade da memória de trabalho—O Teste de Capacidade de Leitura [Important Aspects in the construction, application and interpretation of results of an instrument to measure working memory capacity – the Reading Span Test]. In R. Guaresi & V. W. Pereira (Eds.), *Leitura e Escrita em Avaliação: A ciência em busca de maior esclarecimento da linguagem verbal* (pp. 50-78). Fonema e Grafema.

Tran, P., Carrillo, R., & Subrahmanyam, K. (2013). Effects of online multitasking on reading comprehension of expository text. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *7*(3). https://doi.org/10.5817/CP2013-3-2

Vasilev, M. R., Kirkby, J. A., & Angele, B. (2018). Auditory Distraction During Reading: A Bayesian Meta-Analysis of a Continuing Controversy. *Perspectives on Psychological Science*, *13*(5), 567-597. https://doi.org/10.1177/1745691617747398

Wen, Z. (2015). Working Memory in Second Language Acquisition and Processing: The Phonological/ Executive Model. In Z. Wen, M. B. Mota, & A. McNeill (Eds.), *Working Memory in Second Language Acquisition and Processing* (pp. 41-62). Multilingual Matters. https://doi.org/10.21832/9781783093595-007

Wen, Z. (2016). *Working Memory and Second Language Learning: Towards an Integrated Approach*. Multilingual Matters. https://doi.org/doi:10.21832/9781783095735

Wen, Z., & Li, S. (2019). Working Memory in L2 Learning and Processing. In J. W. Schwieter & A. Benati (Eds.), *The Cambridge Handbook of Language Learning* (1st ed., pp. 365-389). Cambridge University Press. https://doi.org/10.1017/9781108333603.016

Wen, Z., & Schwieter, J. W. (2022). Toward an Integrated Account of Working Memory and Language. In Z. Wen & J. W. Schwieter (Eds.), *The Cambridge handbook of working memory and language*. Cambridge University Press.

Acknowledgments
We acknowledge our fellow colleagues who rated the comprehension questions used in this study. We also acknowledge the Federal Institute of Santa Catarina (*IFSC*) for granting one-year work leave (2021-2022) for the first author to carry out this piece of research.

Funding
This piece of research was funded by the Brazilian National Council for Scientific and Technological Development (*CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico*) in the form of a Scholarship for the first author (GD - 140785/2018-3) and for the third author (PQ2 - 312123/2019-1) and by the Coordination of Superior Level Staff Improvement (*CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior)* in the form of a Scholarship for the third author (PRINT - 88887.584264/2020-00).

Ethics Declarations

Competing Interests

No, there are no conflicting interests.

Rights and Permissions

Open Access