

## Construindo um Teste de Capacidade de Leitura Autoaplicável com apresentação aleatória de estímulos

*Création d'un test de portée de lecture auto-administrable avec présentation de stimuli aléatoires*  
*Construcción de una prueba de lectura autoadministrable con presentación de estímulos aleatorios*  
*Building a Self-Administrable Reading Span Test with Random Stimuli Presentation*

Davi Alves Oliveira<sup>1</sup>, Sidnei Werner Woelfer<sup>2</sup>  
& Lêda Maria Braga Tomitch<sup>2</sup>

<sup>1</sup> Universidade do Estado da Bahia, Jacobina, Bahia, Brasil

<sup>2</sup> Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brasil

### Resumo

O presente artigo relata resultados preliminares de um estudo piloto exploratório que propõe uma versão do Teste de Capacidade de Leitura (Reading Span Test, Daneman & Carpenter, 1980) que não necessite de leitura em voz alta e que possa ser aplicado sem intervenção constante do(a) pesquisador(a), permitindo auto aplicação e viabilizando aplicação em grupos. Além disso, a versão proposta apresenta as sentenças-estímulo em ordem aleatória, de forma a tentar reduzir o efeito de retestagem (Scharfen, Jansen, & Holling, 2018). O instrumento foi desenvolvido em JavaScript com a biblioteca jsPsych (de Leeuw, 2015; de Leeuw & Motz, 2016), podendo ser disponibilizado na web. Para testar a validade da versão proposta, dez participantes foram convidados a realizar o teste na versão em língua portuguesa desenvolvida por Tomitch (2003) que tem as características básicas do teste original, e na versão autoaplicável proposta. Os resultados mostram uma correlação positiva moderada estatisticamente significativa entre as duas versões, apontando para a validade da versão proposta. Adicionalmente, a versão autoaplicável mostrou-se significativamente mais fácil que a versão clássica. Por conta disso, uma limitação no tempo de leitura foi adicionada de forma a aproximar a versão com leitura silenciosa da versão com leitura em voz alta. Verificou-se também que a quantidade de caracteres, quantidade de sílabas, a posição da palavra no nível e o nível em que a palavra é apresentada exercem influência significativa na probabilidade de recordação da palavra a ser recordada. Não foram observados efeitos significativos do número de palavras na sentença lida. Sugere-se, para versões futuras do teste, a seleção de sentenças com palavras finais que tenham a menor variação possível no número de caracteres e quantidade de sílabas ou que tal variação seja levada em consideração na pontuação.

Palavras-chave: Memória de trabalho, Teste de Capacidade de Leitura, testes de memória e aprendizagem, testes de linguagem, JsPsych.

### Resumen

Este artículo reporta resultados preliminares de un estudio piloto exploratorio que propone una versión del Reading Ability Test (Reading Span Test, Daneman & Carpenter, 1980) que no requiere lectura en voz alta y que se puede aplicar sin la intervención constante del investigador, lo que permite la autoadministración y su aplicación en grupos. Además, la versión propuesta presenta las oraciones estímulo en orden aleatorio, con el fin de intentar reducir el efecto de la reevaluación (Scharfen, Jansen, & Holling, 2018). El instrumento fue desarrollado en JavaScript con la biblioteca jsPsych (por Leeuw, 2015; por Leeuw & Motz, 2016) y lo que genera que el mismo pueda estar disponible en la web. Para probar la validez de la versión propuesta, se invitó a diez participantes a realizar la prueba en la versión en lengua portuguesa desarrollada por Tomitch (2003) que tiene las características básicas de la prueba original y, además, la versión autoadministrada propuesta. Los resultados mostraron una correlación positiva moderada estadísticamente significativa entre las dos versiones, lo que apoya la validez de la versión propuesta. Además, la versión autoadministrable resultó ser significativamente más fácil que la versión clásica. Por ello, se añadió una limitación en el tiempo de lectura para acercar la versión con lectura silenciosa a la versión con lectura en voz alta. También se encontró que el número de caracteres, el número de sílabas, la posición de la palabra en el nivel y el nivel en el que se presenta la palabra influyeron significativamente en la probabilidad de recordar la palabra blanco. No hubo efectos significativos del número de palabras en la oración leída. Se sugiere, para futuras versiones de la prueba, la selección de oraciones con palabras finales que tengan la menor variación posible en el número de caracteres y número de sílabas o que dicha variación se tenga en cuenta en la puntuación.

Palabras clave: memoria de trabajo, prueba de capacidad de lectura, pruebas de memoria y aprendizaje, pruebas de lenguaje, JsPsych.

## Résumé

Cet article rapporte les résultats préliminaires d'une étude pilote exploratoire qui propose une version du Reading Span Test (Daneman & Carpenter, 1980) qui ne nécessite pas de lecture à voix haute et qui peut être administrée sans intervention constante du chercheur, permettant l'auto-administration et le groupe administration. De plus, la version proposée présente les stimuli dans un ordre aléatoire, dans le but de réduire l'effet de retest (Scharfen et al., 2018). L'instrument a été développé en JavaScript avec la bibliothèque jsPsych (de Leeuw, 2015; de Leeuw & Motz, 2016), ce qui permet sa possible publication en ligne. Pour tester la validité de la version, dix participants ont été invités à effectuer le test dans sa version portugaise brésilienne développée par Tomitch (2003), qui a fondamentalement les mêmes caractéristiques que la version originale, et dans la version auto-administrable proposée. Les résultats montrent une corrélation positive modérée statistiquement significative entre les deux versions, démontrant la validité de la version proposée. De plus, la version auto-administrable était nettement plus facile que la version classique. En raison de ce résultat, une limitation du temps de lecture a été ajoutée pour augmenter la similitude entre la version de lecture silencieuse et celle de lecture à voix haute. Des effets significatifs ont également été observés concernant le nombre de caractères, le nombre de syllabes, la position du mot au niveau et au niveau de présentation du mot sur la probabilité de rappel des mots à retenir. Aucun effet n'a été observé sur le nombre de mots dans la phrase. Les suggestions pour les futures versions du test incluent la sélection de phrases dont les derniers mots présentent le moins de variation possible dans leur nombre de caractères et de syllabes ou que la méthode de notation tient compte de cette variation.

Mots clés: mémoire de travail, test de durée de lecture, tests de mémoire et d'apprentissage, tests de langue, JsPsych.

## Abstract

This article reports preliminary results of an exploratory pilot study that proposes a version of the Reading Span Test (Daneman & Carpenter, 1980) that does not require reading aloud and that can be administered without constant intervention from the researcher, enabling self-administration and group administration. Furthermore, the proposed version presents the stimuli in random order, aiming at reducing the retest effect (Scharfen et al., 2018). The instrument was developed in JavaScript with the jsPsych library (de Leeuw, 2015; de Leeuw & Motz, 2016), which enables its possible online publication. To test the version's validity, ten participants were invited to perform the test in its Brazilian Portuguese version developed by Tomitch (2003), which basically has the same characteristics of the original version, and in the proposed self-administrable version. The results show a statistically significant moderate positive correlation between the two versions, evidencing the validity of the proposed version. Additionally, the self-administrable version was significantly easier than the classic version. Because of this result, a limitation in the reading time was added to increase the similarity between the silent reading version and the reading aloud one. Significant effects were also observed concerning the number of characters, number of syllables, position of the word at the level and level in which the word is presented on the probability of recall of to-be-remembered words. No effects were observed regarding the number of words in the sentence. Suggestions for future versions of the test include the selection of sentences whose final words have the least possible variation in their number of characters and syllables or that the scoring method accounts for such variation.

Keywords: Working memory, Reading Span Test, memory and learning tests., language tests, JsPsych.

## Introdução

Na busca por modelar a cognição humana, um construto de grande importância é a memória de trabalho (Baddeley & Hitch, 1974), sistema responsável pelo armazenamento temporário e processamento simultâneo de informações provenientes de estímulos internos e externos (Baddeley, 2012; Cowan, 1999; Miyake & Shah, 1999). Um consenso existente a respeito do sistema é sua limitação multifacetada (ver Miyake & Shah, 1999, p. 447-448). Tal limitação varia entre indivíduos e os esforços de muitas pesquisas na área têm sido em explorar, analisar e explicar como tal variação relaciona-se com variações de desempenho em diversas tarefas que envolvem processos cognitivos complexos, como a leitura (ver meta-análises Daneman & Merikle, 1996; Linck, Osthus, Koeth, & Bunting, 2014). Esse tipo de investigação envolve o uso de instrumentos que possam medir a capacidade de memória de trabalho (MT) e, dada a natureza multifacetada de suas limitações e o caráter multimodal do sistema (Baddeley & Logie, 1999; ver também Miyake & Shah, 1999, p. 448-449), é necessário a escolha de instrumentos apropriados que deem conta de medir as diferenças individuais de acordo com os processos específicos da tarefa cognitiva em questão. Se tratando da leitura, um instrumento de mensuração da capacidade da MT é o Teste de Capacidade de Leitura (*Reading Span Test*) desenvolvido por

Daneman e Carpenter (1980), que mede a capacidade do(a) leitor(a) de recordar palavras-alvo enquanto processa sentenças.

O Teste de Capacidade de Leitura (TCL) de Daneman e Carpenter consiste, resumidamente, na leitura em voz alta de sessenta sentenças, mostradas uma a uma. A cada sentença, o(a) participante deve tentar reter a última palavra. As sentenças são organizadas em conjuntos e ao final de cada conjunto é solicitado que o(a) participante recorde as palavras finais das sentenças na ordem em que foram apresentadas. A capacidade de memória de trabalho é considerada pelo nível em que o(a) participante falha em recordar corretamente as palavras ou pelo número total de palavras recordadas corretamente. Uma explicação mais detalhada do teste é dada na seção de método do artigo.

Além de sua versão original em inglês, existem versões do TCL em diversas línguas, como alemão (Carroll et al., 2015; Van Den Noort, Bosch, Haverkort, & Hugdahl, 2008), holandês (Keijzer, 2013; Van Den Noort et al., 2008), norueguês (Van Den Noort et al., 2008), e, mais relevante para o presente estudo, a versão construída por Tomitch (2003) para o português brasileiro, com base no original de Daneman e Carpenter (1980), assim como suas adaptações posteriores que substituíram a apresentação das sentenças em cartões impressos por apresentações em slides com projetor multimídia (Bailer, 2011), e em apresentações em forma de aplicativo web

(Oliveira, 2016). Esses três últimos estudos citados utilizaram o TCL na investigação de possíveis correlações entre a capacidade de memória de trabalho e três outras variáveis: a percepção de organização textual durante a leitura (Tomitch, 2003), a atenção à forma e ao sentido durante a leitura em inglês como língua estrangeira (Bailer, 2011) e a proficiência de leitura em inglês como língua estrangeira (Oliveira, 2016). Resultados dos três estudos mostram correlações estatisticamente significativas entre pontuações do TCL e as variáveis investigadas.

Apesar do sucesso do teste em medir diferenças individuais na capacidade de MT durante a leitura (Bailer, Tomitch, & D'ely, 2013; Oliveira, 2016; Procailo, 2017; Roscioli, 2017; Tomitch, 2003; Woelfer, 2016), algumas limitações podem ser identificadas se desenhos de pesquisa específicos forem considerados. A leitura em voz alta, utilizada para assegurar que o(a) participante leia a sentença completamente e não apenas a palavra a ser recordada, impossibilita a aplicação do teste para mais de um(a) participante simultaneamente no mesmo local pelo(a) mesmo(a) pesquisador(a), o que pode inviabilizar a aplicação do teste para grandes quantidades de participantes num curto espaço de tempo, estratégia muitas vezes necessária para assegurar resultados estatísticos confiáveis. Além disso, o uso de sentenças iguais sempre na mesma ordem pode ser um problema para pesquisas que necessitem de aplicações subsequentes do teste, a exemplo de estudos sobre treinamento de MT (Foster et al., 2017) e estudos longitudinais, que precisam evitar o efeito de retestagem (ver Scharfen et al., 2018).

O presente artigo relata uma proposta de versão do TCL que tenta reduzir as limitações do teste original. O Teste de Capacidade de Leitura Autoaplicável (TCL-A) tem características semelhantes à versão de Daneman e Carpenter (1980) com as seguintes principais diferenças: assim como o teste usado em Oliveira (2016), o TCL-A foi programado com a linguagem de programação Javascript, com uso da biblioteca jsPsych (de Leeuw, 2015), o que faz com que ele possa ser executado em qualquer navegador web e possa, inclusive, ser disponibilizado online, além de possibilitar o registro preciso de tempos de leitura (de Leeuw & Motz, 2016; Pinet et al., 2017; Reimers & Stewart, 2015); ao invés de utilizar sempre as mesmas sessenta sentenças, a cada vez que é executado o teste seleciona aleatoriamente 60 sentenças de um conjunto de 180, possibilitando a construção de testes diferentes a cada aplicação; a cada vez que é executado o teste ordena as sentenças em ordem aleatória de forma a assegurar que, em caso de retestagem, aplicações subsequentes do teste tenham uma probabilidade muito pequena<sup>1</sup> de serem iguais à primeira aplicação e; para remover a necessidade de leitura em voz alta, o teste é programado de forma que, a cada sentença, o(a) participante deve fazer uma tarefa de escolha lexical, ou seja,

escolher uma dentre duas palavras finais que melhor complete o sentido da sentença.

Os resultados aqui relatados são provenientes de um estudo piloto exploratório que, mais do que testar hipóteses, objetivou gerar hipóteses a serem testadas futuramente e propor um desenho metodológico que pode ser seguido em estudos posteriores. Apesar disso, o presente estudo foi guiado por uma hipótese inicial, formulada antes do levantamento de dados, de que as medidas provenientes do Teste de Capacidade de Leitura (TCL) correlacionam-se positivamente com as medidas provenientes do Teste de Capacidade de Leitura Autoaplicável (TCL-A, Hipótese 1). Durante o levantamento de dados, foi levantada a hipótese de que a leitura silenciosa no TCL-A resulta em menor carga cognitiva, evidenciada pelo maior número de palavras recordadas em comparação com o TCL (Hipótese 2). Por fim, após levantamento de dados, quatro hipóteses adicionais foram criadas, relacionadas a características das sentenças e das palavras finais (palavras-alvo) utilizadas no teste. Foram elas: o tamanho da sentença é um bom preditor da probabilidade de recordação da palavra-alvo (Hipótese 3); o tamanho da palavra-alvo é um bom preditor de sua probabilidade de recordação (Hipótese 4); a posição da palavra-alvo no nível é uma boa preditora de sua probabilidade de recordação (Hipótese 5) e; a posição da palavra-alvo no teste é uma boa preditora de sua probabilidade de recordação (Hipótese 6)<sup>2</sup>. Tais hipóteses são exploradas, mesmo considerando-se todas as limitações de um estudo piloto exploratório, como forma a guiar investigações futuras.

Apesar de o TCL objetivar medir a capacidade da MT, ao se considerar a natureza modular do sistema e suas múltiplas fontes possíveis de limitações, como já mencionado, é válido questionar o quê, de fato, o TCL mede. Nesse sentido, é preciso entender melhor as limitações do sistema e os efeitos de tais limitações no processo de leitura. Miyake e Shah (1999, p. 421) apontam que diferentes modelos de MT pressupõem como fontes de limitações o decaimento de informações, a eficiência no controle de atenção ou a falta de controle dos processos de inibição (Engle, 2018), a capacidade de ativação total disponível (Just & Carpenter, 1992), a eficiência ou velocidade de processamento (Daneman & Carpenter, 1980), os processos de codificação ou recuperação afetados por falta de conhecimento ou habilidade, as interferências advindas do efeito de similaridade (Baddeley & Logie, 1999) e a interação entre componentes e subsistemas. A leitura, a partir da perspectiva do processamento da linguagem, pode ser compreendida como um conjunto de processos cognitivos que, executados de modo eficiente, possibilitam ao leitor a construção de uma representação mental do texto lido (Kintsch & van Dijk, 1978; Linderholm & van den Broek, 2002; Rumelhart, 1977, 2013; Rumelhart & McClelland, 1981; Samuels & Kamil, 1984, 1998; van den Broek, Young, Tzeng, & Linderholm, 1999). Diferenças individuais em capacidade

<sup>1</sup> O número de versões possíveis do teste é dado pelo número de conjuntos possíveis de sentenças selecionadas e sua ordem, que pode ser calculado pelo arranjo  $A_{180,60} = \frac{180!}{(180-60)!}$ , aproximadamente  $3,00 \times 10^{130}$ , com

probabilidade de um teste ser igual ao aplicado anteriormente aproximadamente igual a  $3,330 \times 10^{-131}$ .

<sup>2</sup> Uma discussão detalhada sobre os fatores que devem ser considerados na construção de um TCL pode ser encontrada em Tomitch (no prelo).

de MT tendem a estar associadas a diferenças individuais no desempenho da leitura já que tal capacidade determina a quantidade de elementos textuais que podem ser mentalmente integrados em um dado momento, bem como a agregação dessa integração com informações recuperadas da memória de longo prazo (Just & Carpenter, 1980).

Modelos da MT concordam com a limitação do sistema, mas diferem em relação às fontes de limitação. Com base no Modelo Multicomponente proposto por Baddeley e Hitch (1974) e posteriormente revisado por Baddeley (1992, 2000, 2003, 2012), Baddeley, Gathercole e Papagno (1998) e Baddeley e Logie (1999), a capacidade de MT pode ser compreendida como o resultado das interações dos componentes do sistema, a saber, o Executivo Central, componente atencional de controle, a Alça Fonológica, que processa informações de natureza fonológica, o Esboço Visual-Espacial, que processa informações de natureza visual, espacial e possivelmente cinestésicas e o Buffer Episódico, componente de armazenamento de episódios, ou agrupamentos de informações que podem ser de natureza multimodal. Pressupõe-se que a soma das possíveis interações entre os componentes e as diferentes limitações resulta em diferenças individuais mensuráveis. Seguindo essa lógica, e com base em considerações recentes (Baddeley, 2017), diferenças individuais no montante de recursos disponíveis ao Executivo Central, à Alça Fonológica, e ao Buffer Episódico para executar suas funções podem teoricamente explicar possíveis variações no desempenho no TCL e, conseqüentemente, no TCL-A.

Outro modelo de MT bastante citado em estudos sobre o processamento da linguagem, o Modelo dos Processos Embutidos (*The Embedded-Processes Model of Working Memory*) proposto por Cowan (1988, 1993, 1999, 2008, 2015), conceitua a MT como um conjunto de informações que, devido ao seu nível de ativação, se encontra acessível para o cumprimento de tarefas cognitivas (Adams, Nguyen, & Cowan, 2018; Cowan, 1999, 2014, 2017). Segundo esse modelo, o controle e regulação da MT é voluntariamente coordenado por um Executivo Central, porém difere do Modelo Multicomponente ao não propor módulos adicionais. Nesse modelo, os itens presentes na MT são compreendidos simplesmente como informações da Memória de Longo Prazo (MLP) que, de forma análoga, se encontram em estado de ativação (Cowan, 1999, 2008, 2014). O modelo define ainda que a MT possui capacidade limitada quanto à quantidade de informações que podem ser mantidas no foco da atenção, e quanto à duração da ativação das informações presentes no subconjunto ativado da MLP (Cowan, 2008, 2014).

No que se refere ao processamento da linguagem, Cowan (2010, 2014) explica que a compreensão de sentenças enunciadas oralmente demanda que informações sobre protagonistas, ações e características de objetos se mantenham ativadas simultaneamente para que um todo coerente e significativo possa ser construído pelo ouvinte. É possível assumir que um processo similar ocorre durante a leitura. Mesmo tendo o leitor, em comparação ao ouvinte, um comando maior sobre o processamento dos elementos do insumo verbal (Just & Carpenter, 1980), é possível assumir que as limitações do foco da atenção para abrigar as informações essenciais para

a compreensão de sentenças escritas pode comprometer o desempenho dos indivíduos no cumprimento da tarefa de escolha lexical exigida pelo TCL-A. Prevê-se que o êxito no cumprimento de tal tarefa depende do acesso consciente do indivíduo aos elementos que compõem as sentenças para formar o todo coerente. Possíveis falhas na realização dessa tarefa podem também ser explicadas pela forte atuação do campo semântico na percepção lexical dos indivíduos (Rumelhart, 2013). A não agregação dos elementos essenciais no foco da atenção para construir esse campo semântico pode comprometer a percepção lexical, e por conseqüência, a escolha lexical. Além desses aspectos, a possibilidade de intrusão de informações irrelevantes no foco da atenção, especialmente aquelas com carga emocional (Cowan, 2015), pode também afetar a tarefa de escolha lexical. Essa intrusão, por competir por ativação no foco da atenção, pode pôr em risco a construção do campo semântico das sentenças, já que compete com o insumo verbal que precisa ser processado. Além disso, o decaimento de informações na porção ativada da MLP pode interferir na tarefa de recuperação das palavras finais após a leitura de cada conjunto de sentenças. Desse modo, e de acordo com as premissas desse modelo, diferenças entre os escores obtidos no teste podem ser interpretadas como diferenças individuais de capacidade de memória de trabalho (MT).

Por fim, o Modelo de Controle Atencional (*Controlled Attention Model*, Engle, 2018; Engle, Kane, & Tuholski, 1999; Heitz, Unsworth, & Engle, 2005), por sua vez, propõe que a MT é um sistema unitário composto por representações mentais de diversos formatos que, recuperadas da MLP, se encontram em estado de ativação acima do limiar. Além disso, o modelo define que o sistema é composto por mecanismos que executam e mantêm essa ativação, e por um controle atencional limitado (i.e., processos envolvendo controle cognitivo de recursos atencionais). Segundo Shipstead, Harrison e Engle (2016), esse controle atua na alocação de recursos atencionais que garantem o estado ativo e, portanto, a acessibilidade a informações relevantes para a tarefa sob execução. O modelo propõe ainda que diferenças individuais de capacidade de MT primariamente refletem diferenças individuais de controle atencional (Engle, 2018; Engle et al., 1999; Hambrick & Engle, 2002; Rosen & Engle, 1998). Essas diferenças, segundo o modelo, são especialmente observáveis em situações em que o objetivo da tarefa sob execução bem como as informações a ela inerentes competem com informações irrelevantes pelo foco da atenção (i.e., emoções e pensamentos intrusivos e eventos externos distrativos etc.) (Heitz et al., 2005). Além disso, o modelo prevê que o controle atencional é de livre domínio, ou seja, atua na manutenção ativa de toda e qualquer representação mental independentemente de seu formato (Engle, 2018), o que pode explicar as razões pelas quais a capacidade de MT é comumente associada às limitações na execução de uma larga variedade de tarefas cotidianas. Com base nos princípios desse modelo, é possível prever que o TCL, assim como o TCL-A, por impor ao mecanismo de controle atencional altas demandas de processamento, limite a quantidade residual de recursos para a manutenção de informações. Ou seja, o componente de processamento do teste que envolve a leitura, a compreensão e, no caso do TCL-A, a tarefa de escolha lexical pode taxar o

mecanismo de controle atencional ao ponto de fazer com que as palavras a serem recordadas ao fim de cada conjunto de sentenças sejam esquecidas por falta de ativação.

Em resumo, o TCL-A proposto no presente artigo objetiva medir diferenças individuais na capacidade de MT através da leitura silenciosa e escolha lexical, ou seja, escolha da palavra-final mais apropriada para completar o sentido da frase lida, e, simultaneamente, da recordação das palavras-finais escolhidas (palavras-alvo). Pressupõe-se que o teste envolva processos cognitivos semelhantes aos envolvidos em tarefas de leitura em situações comuns do dia a dia e que a medida represente diferenças individuais das interações das várias fontes de limitações de capacidade da MT. Por fim, o teste pode ser aplicado em grupos, sendo recomendado apenas a observação pelo(a) pesquisador(a) ou supervisor(a) de modo a evitar fatores que interfiram na medida, como anotações feitas pelo participante. A próxima seção detalha as etapas seguidas tanto no desenvolvimento do instrumento quanto na análise preliminar de sua validade.

### Método

O estudo piloto envolveu o desenvolvimento do TCL-A, sua aplicação com participantes que também realizaram o TCL, a comparação das pontuações provenientes dos dois testes e adaptações propostas com base nos resultados.

#### *Comitê de Ética*

O presente estudo é parte de um projeto de pesquisa maior que investiga as relações entre capacidade de MT, proficiência de leitura e desenvolvimento do conhecimento de vocabulário durante a leitura em inglês como língua estrangeira. O projeto de pesquisa foi submetido ao Comitê de Ética em Pesquisa com Seres Humanos da Universidade do Estado da Bahia, CAAE 09155919.8.0000.0057, e aprovado sob número de parecer 3.229.002 em 28 de março de 2019 (consulta disponível no link <http://plataformabrasil.saude.gov.br/>).

#### *Participantes*

Dez estudantes matriculados no curso de Letras Língua Inglesa e Literaturas do Campus IV da Universidade do Estado da Bahia, na cidade de Jacobina, estado da Bahia, Brasil, foram convidados para participar da pesquisa. Os critérios de inclusão foram ser maior de 18 anos e ser discente regularmente matriculado no curso. Após serem informados sobre seus objetivos e procedimentos e assinarem o Termo de Consentimento Livre e Esclarecido, sessões individuais de levantamento de dados foram agendadas. As sessões aconteceram em uma sala de estudos do campus da Universidade e tiveram duração média de 30 minutos. Como o objetivo foi comparar dois testes, em um desenho de pesquisa pareado (ou seja, a comparação pressupõe correlação entre as variáveis por envolver o mesmo grupo de participantes em dois momentos distintos) a idade dos participantes não foi considerada como variável do estudo e, por isso, apenas foi assegurado que todos tivessem mais de 18 anos.

Não foi realizado cálculo amostral para a realização dessa etapa do estudo por se tratar de etapa de estudo piloto. Após a análise dos dados, porém, foi feito o cálculo para estabelecer um número de observações que permita maior robustez estatística, como descrito na seção de discussões.

#### *Instrumentos de Levantamento de Dados*

Os instrumentos de levantamento de dados foram o Teste de Capacidade de Leitura (TCL) e o Teste de Capacidade de Leitura Autoaplicável (TCL-A). A versão do TCL utilizada foi a versão em português brasileiro do teste de Daneman e Carpenter (1980) desenvolvida por Tomitch (2003) e adaptada para a biblioteca jsPsych (de Leeuw, 2015) por Oliveira (2016).

*Teste de capacidade de leitura.* O TCL (Daneman & Carpenter, 1980; Tomitch, 2003) é composto por 60 sentenças, não relacionadas entre si, selecionadas de revistas populares de divulgação científica. O tamanho das sentenças varia entre 13 a 17 palavras que são apresentadas em uma ordem preestabelecida, divididas em cinco níveis. Em cada nível, há três conjuntos de sentenças. O primeiro nível é formado por três conjuntos de duas sentenças, o segundo nível é formado por três conjuntos de três sentenças, aumentando desse modo até o quinto e último nível, formado por três conjuntos de seis sentenças. Em outras palavras, no nível 1, o(a) participante deve recordar 2 palavras ao final de cada conjunto, aumentando em 1 palavra nos níveis subsequentes até o nível 5, em que 6 palavras devem ser recordadas ao final de cada conjunto. Cada sentença é mostrada centralizada na tela do computador individualmente e é solicitado que o(a) participante a leia em voz alta e tente recordar a última palavra. Assim que a leitura é concluída, o(a) pesquisador(a) projeta a próxima sentença, que é mostrada no lugar da anterior. Ao final de cada conjunto, no momento da recordação, são mostrados campos de texto no centro da tela. O(A) participante é solicitado(a) a falar em voz alta as palavras recordadas, na ordem em que apareceram, e o(a) pesquisador(a) deve digitar as palavras nos campos de texto indicados na tela. O código-fonte do instrumento, contendo a lista das sentenças utilizadas, encontra-se disponível na plataforma GitHub, podendo ser acessado pelo link <https://github.com/davi-ao/ReadingSpanTest>.

*Teste de capacidade de leitura autoaplicável com estímulos aleatorizados.* Em contraste ao TCL, o TCL-A foi programado com 180 sentenças selecionadas de um corpus de notícias. A cada vez que o teste é executado, uma seleção aleatória de 60 destas 180 sentenças é feita antes de ser apresentada aos participantes. Espera-se que a mudança das palavras contribua para a redução do efeito de retestagem em estudos com múltiplas aplicações do teste por evitar possíveis efeitos de *priming* resultantes da leitura de sentenças idênticas. O tamanho das sentenças varia entre 9 e 15 palavras e, assim como no TCL, estas são apresentadas em cinco níveis, com três conjuntos, cada conjunto com um número crescente de sentenças, começando em 2 sentenças por conjunto (nível 1) e terminando em 6 sentenças por conjunto (nível 5). Ou seja, as sentenças são aleatorizadas, mas os níveis não, algo que é discutido na seção de discussões. Cada sentença é mostrada ao participante com sua última palavra substituída por dois botões. Em um dos botões encontra-se a palavra final original, que

completa corretamente o sentido da sentença, e no outro uma palavra selecionada aleatoriamente das outras sentenças, que geralmente não completa o sentido da sentença corretamente ou não tem as concordâncias sintáticas apropriadas. Existe a possibilidade de, ao acaso, as duas palavras serem apropriadas para completar o sentido da sentença. No início do teste os(as) participantes são instruídos a, caso isso aconteça, selecionarem a que acharem mais apropriada e tentar recordar sempre a palavra que foi clicada. Nesse caso específico, o que importa é que o(a) participante recorde a mesma palavra que clicou. Porém, esse é um problema que precisa ser revisitado em versões futuras do teste. A posição dos botões é aleatorizada. Assim como no TCL, cada sentença é mostrada centralizada na tela do computador individualmente, mas, em contraste ao TCL, o(a) participante é instruído(a) a ler a sentença silenciosamente e a clicar na palavra mais apropriada para completar seu sentido, o que consiste na escolha lexical. Nesse caso, a próxima sentença só é mostrada quando o(a) participante clica em um dos botões. Ao final de cada conjunto são mostrados campos de texto no centro da tela e o(a) participante deve digitar as palavras na ordem em que apareceram. O código-fonte do instrumento, contendo a lista das sentenças utilizadas, também encontra-se disponível na plataforma GitHub, podendo ser acessado pelo link <https://github.com/davi-ao/RandomReadingSpanTest>. Como as sentenças foram tiradas de um corpus de notícias, o instrumento, em sua versão atual, tem como público-alvo leitores maiores de 18 anos, pois algumas sentenças têm conteúdo que pode ser considerado inapropriado para menores (e.g. “Segundo o policial, entre os mortos estão três meninos e uma menina.”).

O corpus do qual as sentenças foram retiradas foi construído por um dos autores com o uso da biblioteca Scrapy (<https://scrapy.org/>) da linguagem de programação Python. O corpus contém 4071 notícias, totalizando aproximadamente 2 milhões de palavras, do portal de notícias G1 (<https://g1.globo.com/>), selecionado por ser o portal de notícias mais acessado no Brasil no momento da construção do corpus. O documento completo em formato JSON, assim como o código utilizado em sua construção, encontra-se disponível online na plataforma GitHub, podendo ser acessado pelo link <https://github.com/davi-ao/G1Corpus>. Os critérios de seleção das sentenças foram: a quantidade de palavras (9 a 15), a possibilidade de isolamento semântico (i.e. a sentença pode fazer sentido fora de seu contexto) e a ordem em que apareceram no corpus. Foram selecionadas as 180 primeiras sentenças que correspondiam a esse critério.

De forma a manter um nível semelhante de validade ecológica do TCL de Daneman e Carpenter (1980) e possibilitar a auto aplicação do teste, tentando evitar os problemas provenientes da variação de poder de processamento e velocidade de conexão dos computadores utilizados para levantamento de dados, o TCL-A foi desenvolvido utilizando-se da biblioteca Javascript jsPsych (de Leeuw, 2015), e de forma a ser aplicado ainda com algum nível de intervenção direta do(a) pesquisador(a). Com o uso da linguagem Javascript o instrumento é primeiramente carregado completamente no navegador e só então executado, necessitando de conexão com a internet apenas no início e ao

final do teste, para registro dos resultados em banco de dados. Desse modo, problemas decorrentes da variação de velocidade de conexão podem ser evitados, ou pelo menos reduzidos significativamente. A utilização da biblioteca jsPsych, por sua vez, além de facilitar o desenvolvimento do instrumento, possibilita uma maior segurança em relação à precisão do registro de tempos de leitura (ver de Leeuw & Motz, 2016; Pinet et al., 2017; Reimers & Stewart, 2015). A presença de um(a) pesquisador(a) ou supervisor(a) durante a aplicação do teste contrasta com propostas como a de Hicks, Foster e Engle (2016) que compararam mensurações de MT através da web com mensurações em laboratório e concluíram que tais medidas na web são viáveis. Porém, os autores identificaram que um dos desafios de tal abordagem é a possibilidade de os participantes simplesmente anotarem os estímulos a serem recordados na tarefa. Os pesquisadores encontraram como solução o uso do alfabeto Klingon, mas essa solução não é considerada satisfatória para o TCL-A por se afastar de uma situação típica de leitura. Por isso, o TCL-A, apesar de ter sido desenvolvido de forma a poder ser acessado na web por qualquer computador com conexão à internet, tem como situação ideal de aplicação um laboratório de informática em que um(a) pesquisador(a) ou supervisor(a) possa observar os(as) participantes.

*Método de pontuação.* Em suas primeiras versões, o TCL é pontuado considerando-se o último nível em que o(a) participante conseguiu recordar corretamente todas as palavras de pelo menos dois conjuntos de sentenças. Desse modo, os valores possíveis para a capacidade de MT são entre 2 e 6. Esse método de pontuação resulta em uma variável discreta com poucos níveis, o que causa um problema em aproximações de sua distribuição para a distribuição normal, um dos parâmetros de testes estatísticos clássicos amplamente utilizados na literatura, sendo o coeficiente de correlação de Pearson o exemplo mais comumente encontrado. Comparando quatro métodos de pontuação do TCL, Friedman e Miyake (2005) relatam que a utilização do total de palavras recordadas, ao invés da pontuação baseada em nível, resulta em distribuições mais próximas da distribuição normal, o que reduz esse problema e, por isso, foi o método de pontuação escolhido no cálculo dos resultados de ambos os testes no presente estudo.

#### *Procedimentos de Levantamento de Dados*

Após o primeiro contato para esclarecimentos e convite para participação da pesquisa, sessões individuais foram agendadas com cada um(a) dos(as) dez participantes. Durante a sessão, o(a) participante realizou os dois testes (TCL e TCL-A). A ordem de realização dos testes foi contrabalanceada de forma que 5 participantes realizaram primeiro o TCL e depois o TCL-A e os outros 5 realizaram primeiro o TCL-A para depois o TCL. Antes do início de cada teste uma sessão de familiarização foi feita. A sessão de familiarização consistiu em uma versão de treinamento do teste com menos sentenças, e com sentenças não utilizadas na etapa de levantamento de dados, de forma que o(a) participante pôde conhecer o passo-a-passo utilizado durante o levantamento de dados. A sessão poderia ser repetida quantas vezes o(a) participante achasse necessário. Alguns(mas) assim o fizeram, mas nenhum(a) participante realizou a sessão de familiarização mais de duas vezes.

*Procedimentos de Análise de Dados*

A análise de dados foi feita com uso da linguagem de programação e ambiente de análise estatísticas R (R Core Team, 2018) com utilização da interface gráfica RStudio (RStudio Team, 2016). Para a comparação entre os resultados de ambos os testes e análise das Hipóteses 1 e 2 foram analisados o nível de significância estatística da correlação entre as pontuações provenientes das diferentes versões e a diferença entre a média de palavras recordadas. Para a análise de possíveis influências de características das sentenças e palavras-alvo na probabilidade de recordação, um modelo logístico multinível foi construído. O modelo logístico foi construído nessa etapa para testar a viabilidade de seu uso em estudos futuros e propor parâmetros. Durante a construção do modelo foram consideradas como variável resposta o resultado da recordação de cada palavra do TCL-A, ou seja, para cada participante foram consideradas as 60 palavras, cada uma como uma observação, com valor 1 para corretamente recordada e 0 para não corretamente recordada. Como variáveis preditoras (1) a posição da palavra-alvo no teste, (2) o número de palavras a serem recordadas em cada conjunto no nível em que a palavra-alvo se encontrava, (3) a posição da palavra-alvo no conjunto, (4) a quantidade de palavras na sentença da palavra-alvo, (5) a quantidade de caracteres da palavra-alvo e (6) a quantidade de sílabas na palavra-alvo. Os(As) participantes foram considerados(as) como fator de efeitos aleatórios. Como as variáveis 2 e 3, assim como as variáveis 5 e 6, são positivamente correlacionadas, o poder preditor das mesmas foram analisados em modelos separados e apenas aquelas que geraram efeitos estatisticamente significativos foram mantidas no modelo final.

**Resultados**

A Tabela 1 mostra o resumo descritivo das medidas dos dois testes e a Tabela 2 resume os dados relativos às características das palavras-alvo e suas respectivas sentenças. Analisando a diferença entre as médias dos dois testes é possível perceber que o TCL-A foi claramente mais fácil do que o TCL, tendo os(as) participantes recordados aproximadamente 11 palavras a mais no TCL-A em comparação ao TCL. Em relação às características das palavras-alvo, vale notar que apesar da grande variabilidade nos tamanhos das sentenças (9 a 15 palavras) e das palavras-alvo (4 a 14 caracteres e 1 a 7 sílabas), a maioria das sentenças e palavras-alvo têm um tamanho semelhante, como observado nos respectivos desvios-padrão, com apenas algumas se distanciando da média. Também já é possível perceber nas tabelas que as distribuições das variáveis parecem se aproximar da distribuição normal considerando a proximidade de suas médias com suas respectivas medianas. Uma análise mais detalhada dessa característica é feita na próxima seção.

Tabela 1. *Resumo descritivo das medidas dos testes*

	TCL	TCL-A
Média	27,3	38,1
Desvio-padrão	9,14	8,39
Mínimo-Máximo	14-42	24-49
Mediana	26,5	39,5

*Nota.* Tanto a versão clássica do teste (TCL) quanto a versão autoaplicável (TCL-A) foram pontuadas considerando-se o total de palavras corretamente recordadas.

Tabela 2. *Resumo descritivo das características das palavras e suas sentenças*

	Tamanho da sentença	Caracteres de palavras	Sílabas na palavra
Média	11,4	7,82	3,3
DP	1,16	2,24	1,01
Mín-Máx	9-15	4-14	1-7
Mediana	11	8	3

*Nota.* DP= desvio-padrão; Mín-Máx= mínimo-máximo.

*Análise descritiva*

O teste de normalidade de Shapiro-Wilk foi utilizado para avaliar se os resultados do TCL e TCL-A podem ser modelados pressupondo-se uma distribuição normal, de forma a melhor informar a escolha de testes estatísticos apropriados para a comparação entre os dois instrumentos. Vale ressaltar que dado o número reduzido de participantes, tais informações são relatadas aqui como forma de guiar estudos futuros. O teste retornou não significativo tanto para os resultados do TCL ( $W = 0,97$ ;  $p = 0,851$ ) quanto para os resultados do TCL-A ( $W = 0,95$ ;  $p = 0,673$ ) o que indica que não há evidências de que as variáveis não sigam uma distribuição normal. A análise dos gráficos Q-Q (Figura 1) também não resultou em nenhuma evidência de que a distribuição das variáveis não possa ser modelada por uma distribuição normal. Desse modo, considerou-se que a distribuição dos resultados dos dois testes pode ser aproximada para uma distribuição normal. Além da normalidade da distribuição, a homogeneidade de variações também foi verificada com o Teste de Homogeneidade de Variações de Levene. O teste retornou resultado não significativo ( $F(1,18) = 0,08$ ;  $p = 0,787$ ), ou seja, não há evidências de que as variações sejam diferentes. Considerando que as variáveis de interesse podem ser modeladas com uma distribuição normal e que há homogeneidade de variação, testes estatísticos paramétricos foram utilizados na investigação das hipóteses que guiaram o estudo, a saber, o coeficiente de correlação produto-momento de Pearson e o teste t de amostras pareadas. Os resultados são discutidos a seguir.

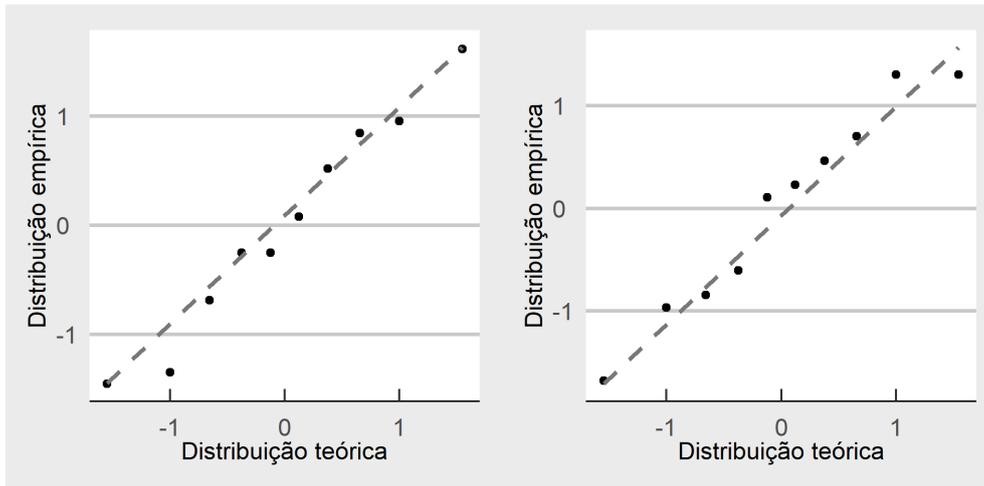


Figura 1. Gráficos Q-Q das medidas do TCL (à esquerda) e do TCL-A (à direita)

*Comparação entre os testes*

A análise dos dados resultou em uma correlação positiva moderada estatisticamente significativa ( $r(8) = 0,57$ ;  $p = 0,043$ ; IC 95% [0,02; 1], unicaudal) mostrada na Figura 2, dando suporte à Hipótese 1. Em relação à Hipótese 2, como

pode ser percebido na Figura 3, a diferença é clara e o teste t pareado mostra que a diferença é estatisticamente significativa ( $t(9) = -4,18$ ;  $p = 0,001$ ; IC 95%  $[-\infty; -6,06]$ , unicaudal). Essa diferença resultou em uma proposta de modificação no teste que será mais bem discutida na seção de discussões.

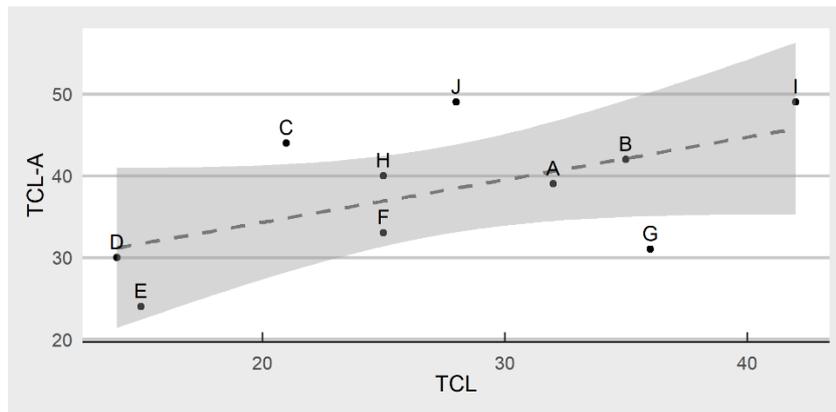


Figura 1. Correlação entre pontuações do Teste de Capacidade de Leitura (TCL) e do Teste de Capacidade de Leitura Autoaplicável (TCL-A)

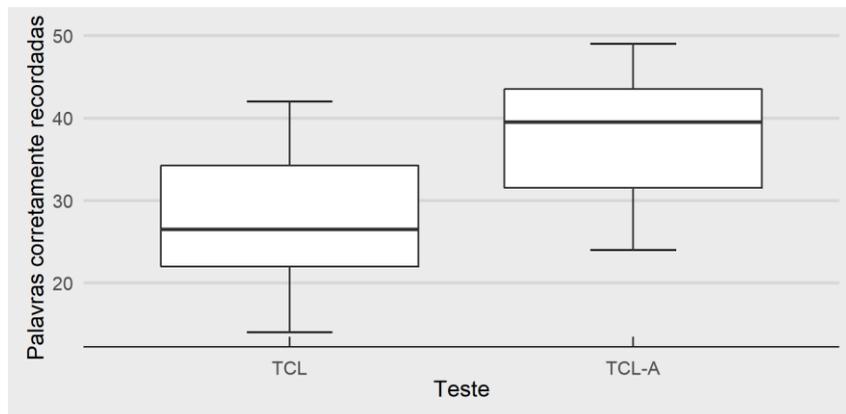


Figura 2. Diagramas de extremos e quartis (boxplots) das pontuações do TCL (à esquerda) e do TCL-A (à direita)

*Modelo logístico multinível*

A possível influência das características das sentenças e palavras finais na probabilidade de recordação é algo já discutido (por exemplo, em Tomitch, no prelo), mas mensurar o nível de tais influências pode melhor informar o controle de tais características ou a consideração destas no cálculo da pontuação do teste. A primeira característica testada foi o tamanho da sentença. Porém, a variável não foi significativa em prever a probabilidade de recordação nem quando considerada isoladamente nem quando associada a outros fatores preditores. Tal resultado pode ser explicado se for considerado que o tamanho da sentença não interfere na probabilidade de recordação de sua palavra-alvo mas sim da palavra-alvo da sentença anterior, já que a manutenção da palavra final na MT ocorre em momento imediatamente posterior ao processamento da sentença em que a palavra-alvo se encontra. Essa explicação precisa ser mais bem investigada em estudos futuros. Por conta da falta de poder preditor, a hipótese não tem suporte dos dados e a variável foi removida do modelo final.

A Hipótese 4 afirma que o tamanho da palavra-alvo é um bom preditor de sua probabilidade de recordação. Espera-se, então, uma associação negativa entre o tamanho da palavra-alvo e sua probabilidade de recordação, ou seja, palavras maiores têm menor probabilidade de recordação em comparação com palavras menores. Duas medidas foram consideradas nessa hipótese: a quantidade de caracteres e a quantidade de sílabas da palavra-alvo. Como as variáveis são positivamente correlacionadas ( $r = 0,90$ ) foi verificado inicialmente quais das duas, consideradas isoladamente, tinham um maior nível de significância. Porém, consideradas isoladamente, nenhuma retornou resultados significativos. Ao considerá-las com outras variáveis em dois modelos separadamente, ambas retornaram resultados significativos. Desse modo, as duas foram mantidas para análise posterior no modelo final.

A Hipótese 5 afirma que a posição da palavra-alvo no grupo é uma boa preditora da sua probabilidade de recordação. Essa hipótese surgiu da observação de que as primeiras e últimas palavras-alvo de cada conjunto tem uma probabilidade de recordação maior do que as palavras-alvo que aparecem nas outras posições. Inicialmente, a variável considerada foi a posição da palavra-alvo no conjunto, variando de 1 a 6, sendo 1 a posição inicial, ou seja, a palavra-alvo que é apresentada primeiro, 2 a segunda posição, ou seja, a segunda palavra-alvo apresentada, até 6, sendo a última posição possível de uma palavra-alvo aparecer, no quinto nível. A variável foi significativa quando considerada isoladamente, porém não se mostrou apropriada para analisar a relação prevista na hipótese pois não resulta em uma relação linear. Isso acontece porque os dados mostram que a probabilidade de recordação começa alta, para a primeira palavra-alvo, reduz até o meio do nível e aumenta novamente na palavra-alvo final. Por exemplo, nos conjuntos do quarto nível, com 5 palavras-alvo, as palavras na

posição 1 têm grande probabilidade de recordação enquanto na posição 2 a probabilidade é menor, reduzindo ainda mais na posição 3, sendo a posição de menor probabilidade de recordação nesse nível. A probabilidade de recordação volta a aumentar na posição 4 e as palavras-alvo na posição 5, por fim, têm grande probabilidade de recordação novamente. Essa tendência pode ser explicada por um efeito de primazia associado a um efeito de recência<sup>3</sup> (ver Freebody & Anderson, 1986, investigação de tais efeitos na recordação de textos). De forma a dar conta de tal observação, a variável “posição no conjunto” foi substituída pela variável “posição relativa” considerando-a como a distância entre a palavra-alvo e a primeira ou última palavra-alvo do nível (considerando a mais próxima). Assim, as primeiras e últimas palavras-alvo de cada conjunto recebem valor 0 nessa variável, enquanto as segundas e penúltimas palavras recebem valor 1 e as terceiras e antepenúltimas palavras recebem valor 2. Quando consideradas juntas, apenas a variável “posição relativa” foi estatisticamente significativa. Desse modo, apenas essa variável foi considerada no modelo final, substituindo-se a variável “posição no conjunto” pela proposta posteriormente.

Por fim, considerando um possível efeito do cansaço do(a) participante ao logo do teste, a Hipótese 6 afirma que a posição da palavra-alvo no teste é uma boa preditora de sua probabilidade de recall. Pressupõe-se, desse modo, uma associação negativa entre a posição da palavra-alvo no teste, como um todo, e sua probabilidade de recordação, ou seja, as palavras-alvo nos níveis iniciais são mais facilmente recordadas que as palavras-alvo nos níveis finais. A variável “posição no teste” foi considerada como a posição da palavra-alvo em relação a todas as outras palavras-alvo do teste, variando de 1, primeira palavra-alvo apresentada, a 60, última palavra-alvo apresentada. A variável foi uma preditora estatisticamente significativa considerada isoladamente, ou seja, palavras-alvo mostradas mais ao final do teste tiveram uma menor probabilidade de recordação. Contudo, esse efeito pode ser associado ao aumento da dificuldade dos níveis e não ao cansaço do participante no teste como um todo. Por isso, uma nova variável, “dificuldade do nível”, foi adicionada e seu poder preditor foi comparado à variável “posição no teste”. A dificuldade do nível é representada pela quantidade de palavras-alvo, que deve ser recordada, no nível, variando de 2 a 6. Quando analisadas juntas, apenas a variável “dificuldade do nível” foi considerada estatisticamente significativa. Desse modo, a variável “posição no teste” foi substituída pela variável “dificuldade do nível”.

Em resumo, não foram observadas evidências para suportar as Hipóteses 3 e 6, ou seja, o número de palavras na sentença da palavra-alvo e a posição da palavra-alvo no teste não são bons preditores de sua probabilidade de recordação, enquanto foram observadas evidências para suportar as Hipóteses 4 e 5, ou seja, o tamanho da palavra-alvo medida tanto pelo seu número de sílabas quanto pelo seu número de caracteres é um bom preditor de sua probabilidade de recordação, assim como a posição relativa da palavra no

<sup>3</sup> De acordo com Ashcraft (1994), itens em posições iniciais (efeito de primazia) e em posições finais (efeito de

recência), em uma lista, tem maior probabilidade de serem lembrados logo após sua apresentação.

conjunto, ou seja, a distância da palavra-alvo para a primeira ou última palavra do conjunto. Por fim, a falta de evidências suportando a Hipótese 6 se deu pela adição de um fator de melhor poder preditivo, a dificuldade do nível, representado pelo número de palavras-alvo a serem recordadas. Como as duas variáveis de medida de tamanho da palavra-alvo (quantidade de caracteres e quantidade de sílabas) são preditoras estatisticamente significativas da probabilidade de recall, dois modelos finais foram construídos ao invés de um. Os dois modelos têm como variáveis preditoras (1) a posição relativa da palavra-alvo em relação às palavras iniciais e finais, (2) a quantidade de palavras a serem recordadas no nível (dificuldade do nível), e o (3) tamanho da palavra-alvo, sendo em um modelo utilizado o número de caracteres e no outro o número de sílabas.

A Figura 4 mostra o efeito estimado da posição relativa da palavra-alvo no conjunto em sua probabilidade de recordação ( $z = -2,65; p = 0,008$ ). A estimativa é que, mantendo-se outros fatores constantes, palavras-alvo apresentadas nas posições iniciais e finais de cada conjunto

(posição 0) têm 70% de probabilidade de recordação enquanto aquelas que aparecem na terceira ou antepenúltima posição (posição 2) tem a probabilidade de recordação reduzida para 53%. A figura mostra também o efeito também estatisticamente significativo da variável “dificuldade do nível” ( $z = -4,02; p < 0,001$ ). As probabilidades estimadas são de 80% no primeiro nível, com 2 palavras-alvo a serem recordadas, reduzindo para 53% no último nível, com 6 palavras-alvo a serem recordadas. Mostra, ainda, o efeito da variável “número de caracteres” da palavra-alvo ( $z = -2,70; p = 0,007$ ), com uma redução de 74% de probabilidade de recordação nas palavras-alvo com 4 caracteres para 48% nas palavras-alvo com 14 caracteres. Por fim, a figura mostra o efeito da variável “número de sílabas” quando utilizada no lugar da variável “número de caracteres” da palavra-alvo, no segundo modelo ( $z = -2,42; p = 0,012$ ). A redução na probabilidade de recordação é de 76% nas palavras-alvo com 1 sílaba para 44% nas palavras-alvo com 7 sílabas.

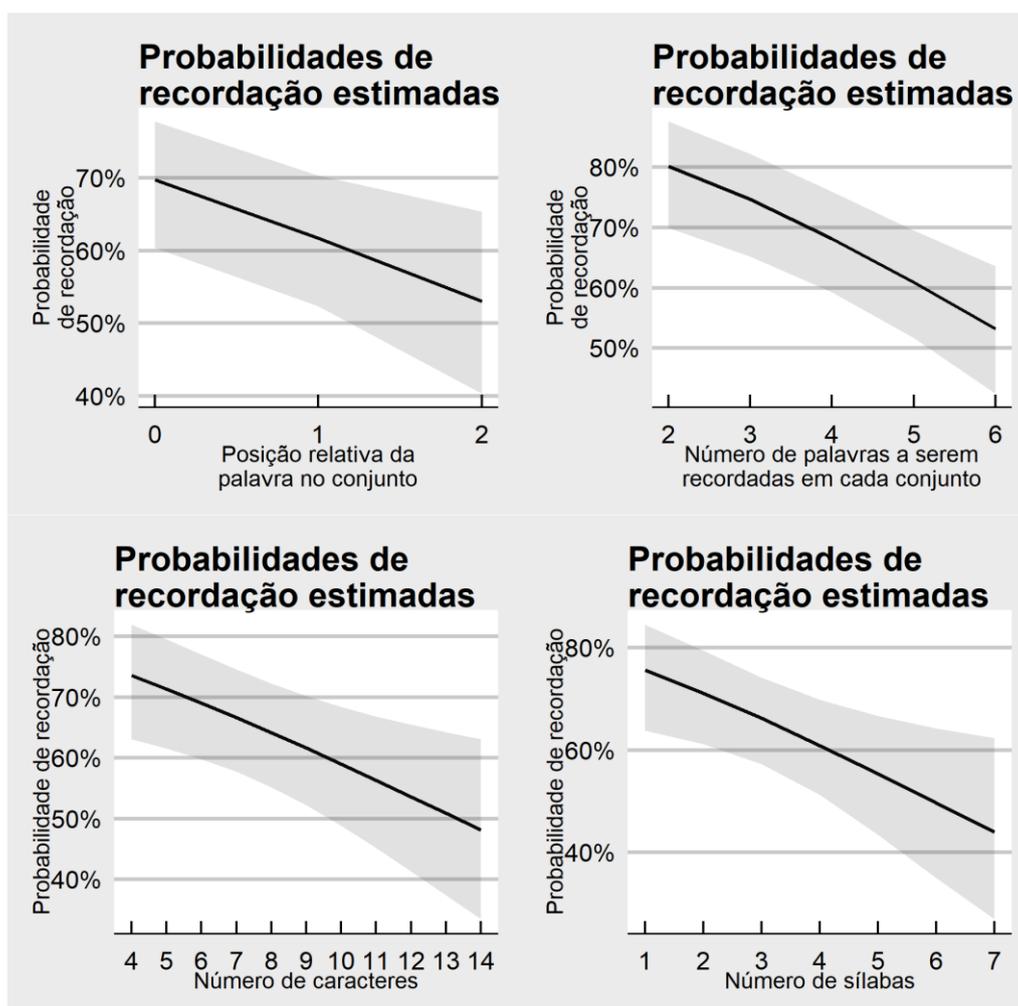


Figura 4. Probabilidades estimadas de recordação das palavras-alvo considerando o efeito de (1) sua posição relativa no conjunto em que é apresentada, (2) do número de palavras-alvo a serem recordadas em cada conjunto no nível em que é apresentada, (3) da quantidade de caracteres na palavra e (4) da quantidade de sílabas da palavra. No primeiro gráfico, o valor 0 representa as posições inicial e final, 1 representa segunda e penúltima posições e 2 representa terceira e antepenúltima posições

### Discussão

Apesar de a Hipótese 1 ter suporte dos dados, é necessário notar as limitações do baixo número de observações que não permite grande poder estatístico, como mostrado pelo intervalo de confiança de 0,02 a 1. Como a Hipótese 2 foi gerada já após dados terem sido parcialmente levantados, o valor de significância do teste da hipótese nula perde o sentido, pois o pressuposto de aleatoriedade da hipótese nula não se mantém. Considerando a natureza exploratória do estudo, os valores de significância foram aqui relatados como forma de simulação para guiar pesquisas futuras.

Com base nos resultados, é possível considerar o TCL-A como instrumento que deve ser melhor investigado como uma opção para medir a capacidade de MT, especialmente em estudos que têm como foco o processo de compreensão leitora. Em relação ao nível de validade comparando-o ao TCL, é possível afirmar que a versão proposta tem o potencial de ter os mesmos níveis observados na versão original, mas por conta da limitação do número de participantes e consequente poder estatístico reduzido, estudos adicionais são necessários para melhor investigar tal afirmação. A principal diferença observada é a relativa facilidade do TCL-A em comparação ao TCL. Tal diferença em dificuldade pode ser problemática, pois diferenças individuais tendem a se mostrar mais evidentes em situações de grande demanda cognitiva (Just & Carpenter, 1992). Vale notar, porém, que nenhum(a) participante conseguiu recordar corretamente todas as palavras. Além disso, o teste foi atualizado para que a sentença seja mostrada apenas por 6,5 segundos de forma a aproximá-lo mais ao tempo de leitura em voz alta da versão original. O tempo foi calculado considerando-se o tempo médio de leitura por palavra, em voz alta, dos participantes do presente estudo durante o TCL. O tempo médio de leitura por palavra foi de 564 milissegundos e como a média de quantidade de palavras na versão atual do TCL-A é de 11,44 palavras, foi sugerido o valor aproximado de 6,5 segundos para cada sentença. Estudos adicionais precisam ser feitos para verificar se tal duração é a mais apropriada e para investigar se a limitação de tempo é suficiente para aproximar o nível de dificuldade do TCL-A ao TCL.

O uso de modelos logísticos na análise dos efeitos das características das palavras-alvo em suas probabilidades de recordação se mostrou de grande valia na análise de quais fatores levar em consideração na escolha de sentenças e palavras-alvo. Deve-se notar, porém, que os resultados aqui expostos têm valor de simulação, já que o número reduzido de participantes não é apropriado para a construção de modelos aleatórios com o número de parâmetros utilizados. Os principais fatores resultantes dos dados do presente estudo foram a quantidade de caracteres e quantidade de sílabas da palavra-alvo, a posição da palavra-alvo no teste e a dificuldade do nível em que a palavra-alvo se encontra. Como a quantidade de caracteres é de processamento computacional mais fácil, sugere-se o uso de tal medida ao invés da quantidade de sílabas,

já que ambos os fatores se mostraram similarmente significativos em prever a probabilidade de recordação, dando-se prioridade a sentenças cujas palavras finais tenham o mesmo número de caracteres. Uma alternativa pode ser utilizada para evitar que as palavras-alvo tenham sempre o mesmo número de caracteres, já que tal medida pode reduzir a validade ecológica do teste. O mesmo método pode ser aplicado em relação à posição da palavra no conjunto. Trata-se de uma proposta, a ser investigada, de se considerar um método de pontuação que atribua diferentes pesos para diferentes palavras-alvo, de acordo com sua posição, número de caracteres e nível. Nesse sentido, a função *predict* do R (R Core Team, 2018) pode ser de grande utilidade. Tal função permite simular resultados controlando os fatores, com base no modelo. Por exemplo, utilizando o modelo do presente estudo, simulando a probabilidade de recordação de palavras de 4 caracteres nas posições inicial ou final em conjuntos do primeiro nível, estima-se uma probabilidade de 88%. Mantendo todos os outros fatores iguais e aumentando a quantidade de caracteres em 1, tem-se a estimativa de 86%. Como a probabilidade de recordação diminuiu 2%, é possível atribuir uma diferença de peso de 2% para cada caractere adicional da palavra-alvo recordada<sup>4</sup>. Assim, por exemplo, enquanto a recordação de uma palavra com 4 caracteres teria peso de 1 ponto, uma palavra com 9 caracteres teria peso de 1,1 ponto no escore final. Como a simulação mostra uma redução de 5% na probabilidade de recordação da palavra de posição 0 (primeira ou última palavra do conjunto) em relação à posição 1 (segunda ou penúltima palavra do conjunto) e de 4% do nível 1 para o nível 2, uma palavra de 4 caracteres na posição 0 no nível 1 teria peso 1 no escore final enquanto uma palavra de 9 caracteres na posição 1 no nível 2 teria peso de 1,19 no escore final.

A sugestão de atribuir diferentes pesos a diferentes palavras-alvo mostra diferenças sutis nos exemplos dados. Porém, o potencial de tal método é utilizar sentenças com alta validade ecológica e ainda assim dar conta das diferenças nas palavras-alvo, que afetam a probabilidade de recordação. Deve-se notar, também, que o presente estudo considerou apenas alguns dos vários fatores que podem afetar a probabilidade de recordação. Fatores como a frequência da palavra-alvo, a familiaridade do participante com a palavra-alvo, e outros fatores relacionados aos níveis semânticos, pragmáticos, fonológicos, prosódicos, assim como aspectos socioculturais dos participantes, podem ter efeitos significativos na probabilidade de recordação, refinando ainda mais a sensibilidade do método de pontuação em relação às diferenças individuais. Como já mencionado anteriormente, um fator de possível influência na probabilidade de recordação é o número de palavras na sentença imediatamente posterior à palavra-alvo, pois é o processamento da sentença lida posteriormente que resulta em processos que podem interferir na manutenção da palavra-alvo na MT.

Em resumo, medidas adicionais podem e devem ser tomadas para refinar o TCL-A. Algumas modificações adicionais, baseadas nos resultados de Friedman e Miyake

<sup>4</sup> O exemplo pressupõe uma relação linear entre as variáveis para facilitar a explicação.

(2005), são a redução do tamanho total do teste, para tornar a execução o mais breve possível, sem afetar a validade e confiabilidade dos resultados, e a aleatorização da ordem dos níveis, de forma a reduzir a previsão do nível de dificuldade destes, o que pode reduzir a ansiedade criada pela expectativa no aumento gradativo da dificuldade nos(as) participantes além de possibilitar momentos de êxito ao longo do teste, e não apenas no início, como apontado pelos autores (p. 589). Essas e outras sugestões são resumidas na próxima seção que também resume as limitações do estudo.

### Considerações Finais

O presente artigo relatou resultados preliminares de um estudo piloto que objetivou testar uma proposta de versão do Teste de Capacidade de Leitura (TCL), chamado aqui de Teste de Capacidade de Leitura Autoaplicável (TCL-A), que não necessita leitura em voz alta e que apresenta um subconjunto aleatório de sentenças em ordem também aleatória a cada vez que é executado, possibilitando autoadministração e uma possível redução no efeito de retestagem. O estudo exploratório conduzido para testar a validade e propor possíveis modificações no teste para avaliações posteriores com número maior de participantes resultou em (1) uma correlação positiva moderada, estatisticamente significativa com resultados do TCL, apontando para sua validade; (2) maior facilidade em comparação com o TCL, e por isso foi proposta uma limitação no tempo de leitura para versões futuras; e (3) a verificação de que a quantidade de caracteres, a quantidade de sílabas, a posição relativa da palavra-alvo no conjunto de sentenças e a dificuldade do nível (quantidade de palavras-alvo a serem recordadas) exercem efeitos significativos na probabilidade de recordação das palavras-alvo. Com base nesse último resultado, sugere-se a criação de novas versões do teste com redução da variabilidade do tamanho da palavra-alvo ou, preferencialmente, um método de pontuação que atribua diferentes pesos a diferentes palavras-alvo considerando – mas não necessariamente limitando-se a – tais características. É preciso reforçar a necessidade de estudos com maior número de participantes para validar os resultados aqui expostos.

#### *Limitações e Sugestões para Pesquisas Futuras*

Apesar da significância estatística da correlação observada, a quantidade reduzida de participantes limitou a possibilidade de maior robustez estatística na avaliação da validade do TCL-A em comparação ao TCL e por isso estudos adicionais com o instrumento são necessários para averiguar os resultados do presente estudo. Além de uma quantidade maior de participantes, estudos futuros também podem se beneficiar da comparação entre métodos de pontuação que atribuem o mesmo peso às palavras recordadas, como é feito tradicionalmente, e métodos de pontuação como o sugerido, que atribui pesos diferentes a diferentes palavras de acordo com as características que afetam sua probabilidade de recordação. Averiguar se o uso de tal método resulta em correlações mais

fortes ou maior poder preditivo com medidas de compreensão de leitura pode informar se as vantagens pressupostas podem ser observadas. Uma sugestão adicional é a análise do efeito de diferentes tempos de exposição das sentenças-estímulo. Com base nos dados do presente estudo, chegou-se ao tempo médio de 6,5 segundos, mas para tornar o instrumento mais confiável seria preciso assegurar menor variabilidade no tamanho das sentenças. Uma sugestão, então, é utilizar tempos de exposição variáveis considerando o número de palavras, proposições, ou ainda caracteres, em cada sentença. Uma diferença adicional entre o TCL e o TCL-A utilizados que não foi levada em consideração no presente estudo é que no TCL é solicitado que o participante fale em voz alta as palavras a serem lembradas para que o pesquisador as digite, enquanto no TCL-A o próprio participante é solicitado a digitar as palavras. Fatores como velocidade de digitação e familiaridade com o layout de teclado utilizado podem influenciar os resultados. Uma sugestão é que versões futuras possam medir a velocidade de digitação (por meio de intervalos entre tecladas, *inter-keystroke-intervals*, por exemplo) para verificar se há algum efeito significativo dessa variável. Uma outra diferença é que ao usar uma seleção de 60 palavras de um conjunto de 180, no TCL-A, cada participante faz o teste com um conjunto diferente de sentenças. Embora tal medida tenha sido tomada para tentar reduzir o efeito de retestagem, o presente estudo não verificou sua eficácia e tal verificação é uma sugestão para estudos futuros. Por fim, o presente instrumento foi desenvolvido de forma a ser facilitado o uso de diferentes listas de sentenças, bastando a modificação de um documento em formato JSON<sup>5</sup>. Desse modo, diferentes corpora podem ser utilizados para extração de sentenças como forma de tentar adaptar o teste para diferentes públicos, considerando as características específicas de cada um, aumentando, assim, a confiabilidade dos resultados em diferentes estudos.

Por fim, após análise do modelo logístico multinível, chegou-se a um modelo com 4 parâmetros, o que permite o cálculo amostral seguindo a “regra de ouro” como descrita por Bujang, Sa’At, Tg Abu Bakar Siki e Lim (2018), que deve ser considerado em estudos futuros. A regra em questão é descrita por  $n = 100 + 50i$ , sendo  $i$  o número de variáveis preditoras. Nesse caso, o número sugerido de participantes seria 300. Porém, no caso do modelo em questão a variável resposta está no nível da palavra, ou seja, há 60 observações por participante e o participante é variável de agrupamento. Nesse caso, com base em Ali et al. (2019), considerando que o número de participantes é o número de grupos de observações, 100 participantes podem resultar em resultados suficientemente robustos. Um fator adicional de grande importância não levado em consideração no presente estudo que deve ser considerado em estudos futuros é o valor de eventos por variável (ver Austin & Steyerberg, 2017).

<sup>5</sup> *JavaScript Object Notation*, formato de armazenamento de dados em texto muito utilizado por ter amplo suporte em várias linguagens de programação.

## Referências

- Adams, E. J., Nguyen, A. T., & Cowan, N. (2018). Theories of Working Memory: Differences in Definition, Degree of Modularity, Role of Attention, and Purpose. *Language, Speech, and Hearing Services in Schools*, 49(3), 340–355. [https://doi.org/10.1044/2018\\_LSHSS-17-0114](https://doi.org/10.1044/2018_LSHSS-17-0114)
- Ali, A., Ali, S., Khan, S. A., Khan, D. M., Abbas, K., Khalil, A., ... Khalil, U. (2019). Sample size issues in multilevel logistic regression models. *PLoS ONE*, 14(11), 1–13. <https://doi.org/10.1371/journal.pone.0225427>
- Austin, P. C., & Steyerberg, E. W. (2017). Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. *Statistical Methods in Medical Research*, 26(2), 796–808. <https://doi.org/10.1177/0962280214558972>
- Baddeley, A. D. (1992). Working Memory. *Science*, 255(2), 556–559. <https://doi.org/10.4249/scholarpedia.3015>
- Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423. [https://doi.org/10.1016/S1364-6613\(00\)01538-2](https://doi.org/10.1016/S1364-6613(00)01538-2)
- Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, 36(3), 189–208. [https://doi.org/10.1016/S0021-9924\(03\)00019-4](https://doi.org/10.1016/S0021-9924(03)00019-4)
- Baddeley, A. D. (2012). Working Memory: theories, models and controversies. *Annu. Rev. Psychol.*, 63, 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., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105(1), 158–173. <https://doi.org/10.1037/0033-295X.105.1.158>
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 47–89). New York: Academic Press.
- Baddeley, A. D., & Logie, R. H. (1999). Working memory: the Multiple-Component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: mechanisms of active maintenance and executive control* (pp. 28–61). Cambridge: Cambridge University Press.
- Bailer, C. (2011). *Working memory capacity and attention to form and meaning in EFL reading* (Universidade Federal de Santa Catarina). Retrieved from <https://repositorio.ufsc.br/bitstream/handle/123456789/95008/299556.pdf>
- Bailer, C., Tomitch, L. M. B., & D'ely, R. C. S. F. (2013). Working memory capacity and attention to form and meaning in EFL reading. *Letras de Hoje*, 48(1), 139–147. Retrieved from <http://revistaseletronicas.pucrs.br/ojs/index.php/fale/article/view/11944/8891>
- Bujang, M. A., Sa'At, N., Tg Abu Bakar Sidik, T. M. I., & Lim, C. J. (2018). Sample size guidelines for logistic regression from observational studies with large population: Emphasis on the accuracy between statistics and parameters based on real life clinical data. *Malaysian Journal of Medical Sciences*, 25(4), 122–130. <https://doi.org/10.21315/mjms2018.25.4.12>
- Carroll, R., Meis, M., Schulte, M., Vormann, M., Kießling, J., & Meister, H. (2015). Development of a German reading span test with dual task design for application in cognitive hearing research. *International Journal of Audiology*, 54(2), 136–141. <https://doi.org/10.3109/14992027.2014.952458>
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, 104(2), 163–191. <https://doi.org/10.1037/0033-2909.104.2.163>
- Cowan, N. (1993). Activation, attention, and short-term memory. *Memory & Cognition*, 21(2), 162–167. <https://doi.org/10.3758/BF03202728>
- Cowan, N. (1999). An Embedded-Processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge: Cambridge University Press.
- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? In W. S. Sossin, J.-C. Lacaille, V. F. Castellucci, & S. Belleville (Eds.), *Essence of Memory* (pp. 323–338). [https://doi.org/10.1016/S0079-6123\(07\)00020-9](https://doi.org/10.1016/S0079-6123(07)00020-9)
- Cowan, N. (2010). The Magical Mystery Four. *Current Directions in Psychological Science*, 19(1), 51–57. <https://doi.org/10.1177/0963721409359277>
- Cowan, N. (2014). Working memory underpins: cognitive development, learning, and education. *Educational Psychology Review*, 26, 197–223. <https://doi.org/10.1007/s10648-013-9246-y>
- Cowan, N. (2015). Second Language Use, Theories of Working Memory and the Vennian Mind. In Z. (Edward) Wen, M. Borges Mota, & A. McNeill (Eds.), *Working Memory in Second Language Acquisition and Processing* (pp. 29–40). <https://doi.org/10.21832/9781783093595-006>
- Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, 24(4), 1158–1170. <https://doi.org/10.3758/s13423-016-1191-6>
- 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](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>
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for

- creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47, 1–12. <https://doi.org/10.3758/s13428-014-0458-y>
- de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods*, 48, 1–12. <https://doi.org/10.3758/s13428-015-0567-2>
- 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>
- Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: mechanisms of active maintenance and executive control* (pp. 102–134). Cambridge: Cambridge University Press.
- Foster, J. L., Harrison, T. L., Hicks, K. L., Draheim, C., Redick, T. S., & Engle, R. W. (2017). Do the effects of working memory training depend on baseline ability level? *Journal of Experimental Psychology: Learning Memory and Cognition*, 43(11), 1677–1689. <https://doi.org/10.1037/xlm0000426>
- Freebody, P. F., & Anderson, R. C. (1986). Serial position and rated importance in the recall of text. *Discourse Processes*, 9(1), 31–36. <https://doi.org/10.1080/01638538609544630>
- 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>
- Hambrick, D. Z., & Engle, R. W. (2002). Effects of Domain Knowledge, Working Memory Capacity, and Age on Cognitive Performance: An Investigation of the Knowledge-Is-Power Hypothesis. *Cognitive Psychology*, 44, 339–387. <https://doi.org/10.1006/cogp.2001.0769>
- Heitz, R. P., Unsworth, N., & Engle, R. W. (2005). Working memory capacity, attention control, and fluid intelligence. *Handbook of Understanding and Measuring Intelligence*, (July), 61–78. <https://doi.org/10.4135/9781452233529.n5>
- Hicks, K. L., Foster, J. L., & Engle, R. W. (2016). Measuring Working Memory Capacity on the Web with the Online Working Memory Lab (the OWL). *Journal of Applied Research in Memory and Cognition*, 5(4), 478–489. <https://doi.org/10.1016/j.jarmac.2016.07.010>
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. In H. Singer & R. B. Rudell (Eds.), *Theoretical models and processes of reading* (3rd ed., pp. 807–839). Newark, DE: International Reading Association.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1), 122–149. <https://doi.org/10.1037/0033-295X.99.1.122>
- Keijzer, M. (2013). Working memory capacity, inhibitory control and the role of L2 proficiency in aging L1 Dutch speakers of near-native L2 English. *Brain Sciences*, 3(3), 1261–1281. <https://doi.org/10.3390/brainsci3031261>
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363–394. <https://doi.org/10.1037/0033-295X.85.5.363>
- 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>
- Miyake, A., & Shah, P. (Eds.). (1999). *Models of Working Memory*. <https://doi.org/10.1017/CBO9781139174909>
- Oliveira, D. A. (2016). *Working Memory Capacity and Mental Translation in EFL Reading Comprehension* (Universidade Federal de Santa Catarina). Retrieved from <https://repositorio.ufsc.br/xmlui/handle/123456789/168075>
- Pinet, S., Zielinski, C., Mathôt, S., Dufau, S., Alario, F.-X., & Longcamp, M. (2017). Measuring sequences of keystrokes with jsPsych: Reliability of response times and interkeystroke intervals. *Behavior Research Methods*, 49, 1163–1176. <https://doi.org/10.3758/s13428-016-0776-3>
- Procailo, L. (2017). *Reading digital texts in L2: working memory capacity, text mode, and reading condition accounting for differences in processes and products of reading* (Universidade Federal de Santa Catarina). Retrieved from <https://repositorio.ufsc.br/handle/123456789/187261>
- R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. Retrieved from <https://www.r-project.org/>
- Reimers, S., & Stewart, N. (2015). Presentation and response timing accuracy in Adobe Flash and HTML5/JavaScript Web experiments. *Behavior Research Methods*, 47, 309–327. <https://doi.org/10.3758/s13428-014-0471-1>
- 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* (Universidade Federal de Santa Catarina). Retrieved from <https://repositorio.ufsc.br/xmlui/handle/123456789/183399>
- Rosen, V. M., & Engle, R. W. (1998). Working Memory Capacity and Suppression. *Journal of Memory and Language*, 39(3), 418–436.

- <https://doi.org/10.1006/jmla.1998.2590>
- RStudio Team. (2016). *RStudio: Integrated Development Environment for R*. Retrieved from <http://www.rstudio.com/>
- Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance VI*. Hillsdale, NJ: Erlbaum.
- Rumelhart, D. E. (2013). Toward an interactive model of reading. In D. E. Alvermann, N. J. Unrau, & R. B. Ruddell (Eds.), *Theoretical models and processes of reading* (6th ed., pp. 719–747). Newark, DE: International Reading Association.
- Rumelhart, D. E., & McClelland, J. L. (1981). Interactive processing through spreading activation. In C. Perfetti & A. Lesgold (Eds.), *Interactive processes in reading* (pp. 37–60). Hillsdale, NJ: Erlbaum.
- Samuels, S. J., & Kamil, M. L. (1984). Models of the reading process. In P. D. Pearson, P. Mosenthal, M. Kamil, & R. Barr (Eds.), *Handbook of Reading Research, vol. III* (pp. 185–224). New York: Longman, Inc.
- Samuels, S. J., & Kamil, M. L. (1998). Models of the reading process. In P. Carrell, J. Devine, & D. Eskey (Eds.), *Interactive Approaches to Second Language Reading* (pp. 22–36). New York: Longman, Inc.
- Scharfen, J., Jansen, K., & Holling, H. (2018). Retest effects in working memory capacity tests: A meta-analysis. *Psychonomic Bulletin & Review*, 25(6), 2175–2199. <https://doi.org/10.3758/s13423-018-1461-6>
- Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working Memory Capacity and Fluid Intelligence: Maintenance and Disengagement. *Perspectives on Psychological Science*, 11(6), 771–799. <https://doi.org/10.1177/1745691616650647>
- Tomitch, L. M. B. (2003). *Reading: text organization perception and working memory capacity*. Florianópolis: UFSC, Departamento de Língua e Literaturas Estrangeiras.
- van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The Landscape Model of Reading: Inferences and the Online Construction of a Memory Representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71–98). Mahwah, NJ: Lawrence Erlbaum Associates.
- Van Den Noort, M., Bosch, P., Haverkort, M., & Hugdahl, K. (2008). A standard computerized version of the reading span test in different languages. *European Journal of Psychological Assessment*, 24(1), 35–42. <https://doi.org/10.1027/1015-5759.24.1.35>
- Woelfer, S. W. (2016). *Constructing Meaning From Cartoons : the effects of EFL reading proficiency and working memory capacity on the processing of verbal and pictorial information* (Universidade Federal de Santa Catarina). Retrieved from <https://repositorio.ufsc.br/xmlui/handle/123456789/167472>