

## Desafios na avaliação da compreensão leitora: demanda cognitiva e leiturabilidade textual

*Défis dans la d'évaluation de la compréhension de la lecture: exigences cognitives et lisibilité textuelle*

*Desafíos en la evaluación de la comprensión lectora: demandas cognitivas y leiturabilidad textual*

*Challenges on reading comprehension assessment: cognitive demands and text readability*

Lucilene Bender de Sousa<sup>1</sup>, & Lilian Cristine Hübner<sup>1,2</sup>

<sup>1</sup> Pontifícia Universidade Católica do Rio Grande do Sul – PUCRS, Brasil

<sup>2</sup> Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brasil

Agradecimento: À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo fomento de pesquisa através do processo BEX 18592/12-0. Como também ao Conselho Nacional de Desenvolvimento Científico e Tecnológico.

### Resumo

Este artigo aborda importantes desafios na construção de tarefas de avaliação da compreensão leitora, revisando dois aspectos cruciais: as demandas cognitivas presentes nas tarefas e a leiturabilidade textual. Ao longo da revisão reflete-se sobre a complexidade envolvida na avaliação da compreensão leitora juntamente com os cuidados e os critérios a serem tomados quando da construção das tarefas em relação aos dois aspectos discutidos (demandas cognitivas e leiturabilidade). Na primeira parte, o artigo apresenta diferentes tipos de tarefas que buscam investigar o produto da compreensão leitora, destacando seus objetivos e limitações. Então, discute os processos cognitivos subjacentes que influenciam no produto da compreensão leitora como o reconhecimento de palavra, o conhecimento prévio, as memórias e a habilidade de expressão verbal. Na segunda parte, aborda métodos quantitativos e qualitativos de análise da leiturabilidade dos textos, bem como os fatores que devem ser levados em consideração quando da escolha do texto para compor as tarefas avaliativas da compreensão leitora. Por fim, conclui que, apesar dos progressos da pesquisa sobre leitura nas últimas décadas, a avaliação da compreensão leitora tem incorporado poucos desses avanços. Há uma lacuna em relação a estudos integrados capazes de promover a interface entre a pesquisa em leitura e a pesquisa em avaliação da compreensão leitora.

Palavras-chave: leitura; compreensão; cognição; leiturabilidade; avaliação.

### Resumen

Este artículo aborda importantes desafíos en la construcción de tareas de evaluación de la comprensión lectora, mediante la revisión de dos aspectos cruciales: las demandas cognitivas presentes en las tareas y la legibilidad textual. Durante la revisión se refleja la complejidad implicada en la evaluación de la comprensión lectora, así como los cuidados y los criterios que se deben tener en cuenta en la elaboración de tareas en relación a los dos aspectos discutidos (demandas cognitivas y legibilidad). En la primera parte, el artículo presenta diferentes tipos de tareas que pretenden investigar el producto de la comprensión lectora, destacando sus objetivos y limitaciones. Luego, se analizan los procesos cognitivos subyacentes que influyen en el producto de la comprensión lectora como el reconocimiento de palabras, el conocimiento previo, las memorias y la habilidad de expresión verbal. En una segunda parte se abordan los métodos cuantitativos y cualitativos de análisis de la legibilidad de los textos y los factores que deben tenerse en cuenta en la selección de los textos para las tareas de evaluación de la comprensión lectora. Finalmente, se concluye que, a pesar de los avances de la investigación sobre lectura en las últimas décadas, la evaluación de la comprensión lectora ha incorporado solo algunos de estos avances. Hay una brecha en relación con la promoción de la interfaz entre los estudios de investigación en la lectura y la investigación sobre la evaluación de la comprensión lectora.

Palabras clave: lectura; comprensión; cognición; legibilidad; evaluación.

### Résumé

Cet article traite des défis clés dans la construction de tâches d'évaluation de la compréhension de la lecture, et il fait l'examen de deux aspects cruciaux: les exigences cognitives présentes dans les tâches et la lisibilité textuelle. Tout au long de l'article, se reflète sur la complexité de l'évaluation de la compréhension de la lecture, les soins et les critères à prendre lors de la construction des tâches pour les deux aspects discutés (exigences cognitives et lisibilité textuelle). Dans la première partie, l'article présente différents types de tâches qui cherchent à enquêter sur le produit de la compréhension en lecture, en soulignant ses objectifs et ses limites.

Artigo recebido: 11/04/2014; Artigo revisado: 04/06/2014; Artigo revisado (2ª revisão): 03/12/2014; Artigo aceito: 17/11/2014.

Correspondências relacionadas com este artigo devem ser enviadas a Lucilene Bender de Sousa, Pontifícia Universidade Católica do Rio Grande do Sul, Faculdade de Letras, Programa de Pós-Graduação em Letras (Linguística), Avenida Ipiranga, 6681, Prédio 8, sala 427 CEP 91530-000, Porto Alegre, RS – Brasil.

E-mail: [lenebender10@gmail.com](mailto:lenebender10@gmail.com)

DOI: 10.5579/rnl.2013.0237

Ensuite, examine des processus sous-jacents à influencer le produit de compréhension de la lecture comme: la reconnaissance des mots, la connaissance préalable, les mémoires et la capacité d'expression verbale. La deuxième partie traite des méthodes quantitatives et qualitatives de l'analyse de la lisibilité des textes et quels facteurs doivent être pris en compte lors de la sélection de texte pour composer les tâches d'évaluation de la compréhension en lecture. Enfin, il conclut que, malgré les progrès des recherches sur la lecture au cours des dernières décennies, le moyen d'accéder à l'évaluation de la compréhension de la lecture ne semble pas avoir changé ou essayé d'intégrer ces avancées. Il y a une lacune dans les études intégrées capables de promouvoir l'interface entre les études de recherche en lecture et de la recherche sur l'évaluation de la compréhension de lecture.

Mots-clés: lecture; compréhension; cognition; lisibilité; evaluation.

## Abstract

This paper addresses major challenges on building tasks for reading comprehension assessment, reviewing two important aspects: the cognitive demands of tasks and text readability. This review reflects on the complexity involved in reading comprehension assessment, the cautions and criteria to be considered during task construction dealing with the two aspects discussed (cognitive demands and readability). In the first part, the article presents different kinds of tasks that assess the product of reading comprehension, highlighting their objectives and limitations. Then, it discusses some hidden cognitive processes that influence on the product of reading comprehension, such as word recognition, world knowledge, memories and verbal ability. In the second part, it focuses on quantitative and qualitative methods of text readability analysis, as well as on which factors should be taken into consideration when choosing a text to be part of a reading assessment task. Finally, it concludes that, despite the progresses in reading research in the last decades, reading comprehension assessment has not incorporated many advances. There is a lack of integrated studies able to promote the interface between reading research and research on reading comprehension assessment.

Keywords: reading; comprehension; cognition; readability; assessment.

## Introdução

*“Vemos um pouco mais da compreensão do que Platão viu da realidade nas sombras da caverna.”*  
(Pearson & Hamm, 2005, p. 14)

Avaliar a compreensão leitora é uma tarefa que apresenta inúmeros desafios devido à sua natureza cognitiva. A compreensão leitora requer que múltiplos processamentos cognitivos funcionem colaborativamente como: reconhecimento de palavras, integração semântica local e global, memória de trabalho, memória de curto e longo prazo, inferência, monitoramento; de forma que, à medida que percorre as linhas do texto, o leitor consiga construir incrementalmente os sentidos gerando um modelo textual e situacional coerente (Kintsch, 1998). Como esses processos ocorrem internamente na mente do leitor, autores como Caldwell (2008) afirmam que não é possível observar o processo de compreensão acontecendo. As pesquisas com neuroimagem são uma tentativa de adentrar no universo cognitivo e examinar o que ocorre quando uma pessoa está lendo, porém, ainda estamos longe de poder descrever as operações cerebrais e os processos cognitivos lá desenvolvidos da forma como descrevemos os demais sistemas fisiológicos corporais. Sendo assim, o que muitos testes na verdade avaliam é o produto da compreensão leitora (Fletcher, 2006), cuja análise depende das lentes interpretativas adotadas pelo avaliador. Sweet (2005) igualmente alerta para a importância da construção de um sistema de avaliação da compreensão leitora confiável, capaz de servir tanto às necessidades da pesquisa quanto do ensino.

Este artigo surgiu dos desafios encontrados ao construir tarefas de avaliação da compreensão leitora. Antes de iniciar a construção foi necessário realizar uma investigação teórica sobre os diversos fatores nela implicados desde os individuais (atenção, memória de trabalho, motivação, etc.) e os textuais/discursivos (extensão textual,

complexidade textual, etc.) até os situacionais (ambiente de leitura, duração da tarefa, complexidade e tipo de instrução, etc.). Deste estudo, foram merecedores de maior atenção as demandas cognitivas presentes nas tarefas e a leiturabilidade textual. Dessa forma, o artigo objetiva abordar esses dois fatores por meio de uma revisão teórica não sistemática e uma análise crítica de estudos que discutem a forma como esses fatores estão implicados nas tarefas de avaliação da compreensão de leitores saudáveis, tendo em vista servir de auxílio aos pesquisadores e clínicos interessados tanto na pesquisa sobre avaliação da compreensão leitora quanto na construção de tarefas para este fim.

O artigo está organizado em duas seções. Na primeira, revisa-se o que estudos apontam como demandas cognitivas presentes nos diferentes tipos de tarefas, com destaque às memórias, ao conhecimento prévio, à habilidade escrita, dentre outros. Na segunda seção, discutem-se os métodos quantitativos e qualitativos de análise da leiturabilidade textual, tendo como foco os critérios para a escolha de textos adequados para a construção das tarefas de avaliação da compreensão leitora. A revisão apresenta diversos aspectos a serem considerados na análise da leiturabilidade que vão desde o nível da palavra, em termos linguísticos, até considerações sobre o conhecimento prévio e a motivação do leitor.

## *Demandas cognitivas decorrentes de parâmetros psicométricos*

Apesar da existência de diversas formas de avaliar a compreensão leitora, nem todas parecem examinar o mesmo construto: algumas avaliam o produto da compreensão e não o processo; é o caso de tarefas que requerem a resposta a questões abertas ou de múltipla escolha que se diferenciam de estudos como ERP (Event Related Potentials - Potenciais relacionados a eventos) ou fMRI (Functional Magnetic Resonance Imaging – Imagem por ressonância magnética funcional) que buscam verificar o processo de compreensão no momento em que ele ocorre. Outras avaliam mais os

processos subjacentes do que a compreensão propriamente dita (Nation & Snowling, 1997), tarefas de resumo, por exemplo, dependem em grande parte da lembrança da base textual e da habilidade de expressão escrita do leitor. Inicialmente, é preciso ressaltar que, de fato, o que elas avaliam são amostras de desempenho e comportamentos que julgamos serem efetivos para a compreensão. Entre os principais problemas estão a validade e a confiabilidade dos resultados. O primeiro consiste em analisar se o teste realmente avalia o que pretende avaliar, ou seja, se é uma forma eficiente de avaliação. O segundo refere-se à precisão dos resultados, considerando a existência de possíveis erros (Nunnally, 1959). Nenhum teste é preciso de modo a abranger todos os aspectos a serem avaliados, por isso, análises estatísticas são necessárias para buscar atenuar suas limitações por meio de medidas do nível de acurácia e estimativa de erros e, em alguns casos, de tempo para execução das tarefas. Vejamos brevemente alguns tipos de tarefas que avaliam o produto da compreensão.

Medidas que verificam a compreensão por meio de seu produto podem ser classificadas em dois grupos: receptivas e produtivas (Caldwell, 2008). A Tabela 1 apresenta esses dois grupos, destacando suas principais limitações, as quais serão discutidas na sequência.

Tabela 1. *Tipos de tarefas de avaliação da compreensão leitora e suas limitações*

Tipo de tarefa	Grupo	Limitações
<b>Questões de múltipla escolha</b>	Receptiva	Sofre efeitos de processos de adivinhação e uso de estratégias de resolução de problemas. Pode levar a inferências não realizadas durante a leitura.
<b>Verdadeiro ou falso</b>	Receptiva	Sofre efeitos de processos de adivinhação e uso de estratégias de resolução de problemas. Pode levar a inferências não realizadas durante a leitura.
<b>Técnica de verificação de sentenças</b>	Receptiva	Sofre efeitos de processos de adivinhação e uso de estratégias de resolução de problemas. Pode levar a inferências não realizadas durante a leitura. Depende em grande parte da memória.
<b>Questões abertas</b>	Produtiva	Subjetividade na correção e interferência da capacidade de expressão escrita do leitor.
<b>Resumo</b>	Produtiva	Subjetividade na correção, interferência da capacidade de expressão escrita do leitor e conhecimento de como produzir o gênero resumo. Grande carga de memória.
<b>Reconto</b>	Produtiva	

		Subjetividade na correção, interferência da capacidade de expressão oral do leitor e grande carga de memória.
<b>Técnica de Cloze</b>	Produtiva	Avalia processos de compreensão locais especialmente conhecimento sintático e vocabulário, menos eficiente na avaliação da compreensão global.

Caldwell (2008) distingue dois grupos de medidas de compreensão: o de questões que têm respostas selecionadas (receptivas) e o de questões que requerem a produção de respostas (produtivas). No primeiro grupo encontram-se as tarefas do tipo verdadeiro ou falso, múltipla escolha e técnica de verificação de sentenças. Essa última é semelhante ao verdadeiro ou falso e consiste em apresentar sentenças para o leitor responder julgando se estavam ou não presentes no texto lido (Carlisle, 1989). São quatro os tipos de sentenças: originais, paráfrases, mudança de significado e distratoras. As tarefas receptivas têm a vantagem de serem de correção fácil e objetiva, embora demandem maior tempo para a construção da tarefa. Já o segundo grupo abarca tarefas do tipo recontar, resumir, responder a questões abertas e preencher lacunas do texto (cloze). Elas não demandam tanto tempo de preparação, todavia a sua correção é laboriosa e têm a desvantagem de ser muitas vezes subjetiva. Uma forma de contornar parcialmente a subjetividade é a avaliação por juízes treinados (Fonseca et al. 2011), estabelecendo-se critérios claros de correção que permitam alto grau de coerência entre as avaliações.

Os dois grupos acima apresentados acessam habilidades diferentes: reconhecimento (grupo de seleção das respostas) e produção (grupo de produção de respostas). Se por um lado reconhecer parece ser mais fácil do que produzir, é preciso considerar a tarefa de análise de cada uma das alternativas, pois o leitor pode compreender o texto, mas não conseguir interpretar as alternativas. Além disso, é possível que o participante utilize de adivinhação e/ou de estratégias compensatórias em algumas respostas, o que diminui a confiabilidade dos resultados, especialmente em testes com poucas alternativas, como no caso de verdadeiro ou falso. Simples estratégias como observar a alternativa que completa gramaticalmente a frase, evitar alternativas que contenham palavras absolutas como “todos, nenhum, sempre, nunca”, eliminar as alternativas menos coerentes, optar por alternativas com maior número de informações (mais completas), evitar mudar sua primeira resposta, entre outras, podem proporcionar um desempenho acima da real capacidade de compreensão do leitor.

A produção de respostas escritas e orais, por sua vez, também envolve demandas externas ao processo de compreensão, como a habilidade de expressão verbal e de memória trabalho, bem como a memória declarativa a qual disponibiliza o conteúdo já conhecido sobre o texto. A dificuldade na correção das respostas escritas ocorre devido a respostas mal elaboradas, incompletas e cópias do texto. Por isso, é recomendado que se estabeleçam previamente critérios claros de correção e, de preferência, que seja feita a

apreciação das respostas por mais de um avaliador, ou seja, a análise de juízes (Fonseca et al. 2011). Hoje existem recursos tecnológicos que podem ajudar nas correções de questões. A maioria deles é desenvolvida para instituições de ensino e professores havendo inclusive a opção de correção automática para questões de múltipla escolha e geração de relatórios de desempenho. No entanto, a correção de questões abertas é ainda um desafio. Há tentativas de facilitar o processo de avaliação como o software Powergrading ([http://research.microsoft.com/enus/um/people/sumitb/gradin\\_g/](http://research.microsoft.com/enus/um/people/sumitb/gradin_g/)) que, por meio do aprendizado de máquinas e processamento de linguagem natural, identifica e agrupa padrões de respostas. O programa permite que o avaliador explore os grupos e subgrupos, classificando as respostas em correta, parcialmente correta ou incorreta, e forneça feedback aos alunos. Segundo uma pesquisa realizada por Brooks, Basu, Jacobs e Vanderwende (2014), criadores do sistema, com um clique o professor é capaz de corrigir um grande número de repostas abertas sem perder a eficiência no seu julgamento.

Outra tarefa produtiva comumente utilizada é o teste cloze que consiste no apagamento de palavras do texto, tradicionalmente uma a cada cinco palavras, se solicitando ao leitor que complete as lacunas com a palavra original. Algumas versões adaptadas do teste apagam apenas uma determinada categoria de palavras e são mais flexíveis nas respostas, aceitando como corretas palavras sinônimas ou de sentido bem aproximado ao do texto original. Apesar de apresentar grande correlação com testes de compreensão leitora, há dúvidas sobre o que exatamente ele mede (Farr & Carrey, 1986): vocabulário, sintaxe e/ou compreensão. O cloze também é empregado no ensino e na verificação da leiturabilidade de textos, podendo avaliar tanto o produto, quando respondido por escrito, ou o processo da leitura, quando respondido oralmente e concomitante à leitura. Apesar de parecer de fácil preparação e pontuação, é preciso realizar uma pesquisa bibliográfica detalhada antes de construir e pontuar o teste, pois existem diversos procedimentos para isso. Conforme Kobayashi (2009), variações podem ocorrer na leiturabilidade do texto, no método de pontuação, no ponto inicial e na taxa de apagamento. A pontuação, por exemplo, pode considerar como correta apenas uma palavra ou outras palavras que o avaliador julgue aceitáveis; no entanto, ao assumirem-se pontuações mais flexíveis pode-se correr o risco de acessar outros processos que não a compreensão leitora, como a habilidade linguística geral. A interpretação do resultado obtido através do cloze pode trazer desafios para o avaliador. Autores como Wagner (1986) observam que nem sempre há consenso sobre os critérios de interpretação da pontuação do cloze para o estabelecimento de níveis de compreensão leitora.

Há processos subjacentes -“hidden processes” - (Cain & Oakhill, 2006) que muitas vezes são desconsiderados ou nem mesmo percebidos pelos pesquisadores porque ocorrem simultaneamente à leitura e variam conforme a tarefa e seu nível de dificuldade. Entre eles estão: reconhecimento de palavra, habilidade de produção oral e escrita, resolução de problemas, predição e memórias. Um exemplo é a avaliação da compreensão por meio de questões referentes unicamente

ao conteúdo explícito textual. Esse tipo de questão requer apenas a localização de informações no texto, avaliando, assim, apenas a capacidade de reconhecimento das palavras e não a de compreensão leitora. Portanto, a não observação de processos subjacentes pode levar a uma avaliação imprecisa da compreensão leitora, podendo superestimar ou subestimar a real capacidade do leitor.

Cain e Oakhill (2006) examinaram a demanda cognitiva presente em várias tarefas e concluíram que, por exemplo, os testes cloze dependem muito do reconhecimento de palavras, avaliando mais a compreensão local do texto, logo, não são apropriados para a verificação da compreensão global. Elas também esclareceram que verdadeiro ou falso e múltipla escolha não são bons indicadores de inferências, uma vez que, por apresentarem as respostas prontas, não garantem que o processo tenha sido realizado durante a leitura. As autoras sugerem o uso de questões abertas para o diagnóstico de dificuldades de compreensão leitora por possibilitarem a análise dos padrões de erros (Oakhill, Cain, & Elbro, 2014) e, assim, a identificação dos tipos de dificuldades.

Há muito se debate a existência de diferenças em termos de eficácia entre tarefas avaliativas que empregam questões abertas e questões de múltipla escolha (Chopping & Purves, 1969; Bennett & Ward, 1993; Ozuru, Briner, Kurby & McNamarra, 2013). Em termos de processos, elas requerem estratégias diferenciadas. A resolução de questões abertas parece ser mais difícil porque requer a produção, e não o reconhecimento das respostas. No entanto, esse mesmo aspecto, que parece favorável, resulta em demandas linguísticas e cognitivas extras, já que o desempenho do leitor poderá variar conforme a sua habilidade em produzir as respostas oralmente, ou por escrito, bem como conforme habilidades cognitivas como síntese e capacidade de inibição de informações irrelevantes. Por conseguinte, leitores saudáveis com dificuldades de expressão verbal ou em habilidades cognitivas relacionadas à leitura e à escrita podem ficar em desvantagem. Há estudantes que, apesar de terem uma excelente compreensão, sofrem para se expressar ou redigir as suas respostas.

As questões de múltipla escolha também apresentam processos subjacentes, por exemplo, o reconhecimento da resposta correta pode ser feito por meio de estratégias de resolução de problemas, como o método de eliminação, em que se descartam as alternativas menos prováveis; além disso, mesmo que o leitor não saiba a resposta, ele sempre tem a opção de tentar adivinhar a alternativa correta, podendo escolher aquela que nem havia cogitado durante a leitura, mas que lhe parece plausível e a mais correta dentre as propostas. A participação de processos de memória de trabalho e memória episódica também se fazem necessários para que o leitor possa comparar o conteúdo proposicional do texto ou da base textual com o das alternativas, buscando a que mais se assemelha à representação textual por ele construída. Com alguns leitores ocorre o contrário do exemplo anteriormente dado, eles sabem explicar as respostas com as suas próprias palavras, mas não conseguem identificar e selecionar a alternativa correta.

Campbell (2005) investigou a diferença entre essas duas formas de avaliação. Em uma breve revisão, ele observou que os estudos que buscaram comparar questões

abertas e de múltipla escolha (Langer, 1987; Campbell, 1999) seja por meio de métodos estatísticos ou protocolos verbais, são inconclusivos, não havendo consenso sobre qual seja a melhor forma. O autor afirma que, apesar das críticas às suas limitações, numerosos estudos demonstram a validade e utilidade dos testes de múltipla escolha (van den Bergh, 1990; Berk, 1996). Em sua própria investigação, o pesquisador decidiu estudar por meio de protocolos verbais os processos e estratégias utilizadas pelos leitores na resolução das duas formas de questões. O participante com melhor desempenho em compreensão leitora não mostrou diferença de processamento em questões de múltipla escolha e abertas, enquanto que o participante com desempenho médio mostrou maior engajamento na resolução de questões abertas, respondendo de forma menos cuidadosa às questões de múltipla escolha. Segundo Campbell (2005), essa diferença pode ser resultado não só do formato das respostas, mas também dos textos e das características individuais de cada leitor, como proficiência em leitura, conhecimento de mundo, experiência na realização de determinados tipos de tarefas e na leitura de determinados gêneros textuais, interesse e motivação pelos textos, etc. Questões abertas parecem ser mais adequadas aos leitores com dificuldade, pois requerem maior engajamento na resolução da tarefa, visto que as questões não podem ser resolvidas por meio de uma simples escolha aleatória, pois mesmo com dificuldade, o leitor precisa produzir alguma resposta. O resultado de Campbell (2005) está em conformidade com o do estudo de Nation e Snowling (1997) que, ao compararem vários testes de compreensão leitora, concluíram que avaliações individuais com questões abertas são as mais adequadas quando o objetivo é identificar leitores com dificuldades de compreensão, apontando, igualmente, para a importância de se considerarem as características individuais desses leitores.

Diferenças de demandas cognitivas podem surgir devido ao tema do texto e ao tipo de pergunta elaborada. Priebe, Keenan e Miller (2011) encontraram indícios de que o conhecimento prévio acerca do tema textual facilita tanto o reconhecimento de palavras quanto o processo de compreensão em leitores com baixo desempenho, que parecem utilizar seus conhecimentos como estratégia de compensação para dificuldades de compreensão leitora. A facilitação também pode ocorrer ao responder a certos tipos de questões como as estudadas por Keenan e Betjemann (2006). Os pesquisadores colocaram à prova o teste GORT (Gray Oral Reading Test, de Wiederholt & Bryant, 2001), com o objetivo de avaliar suas características psicométricas. Eles descobriram que muitas de suas questões podiam ser respondidas sem a leitura dos textos. Ao aplicar inicialmente o teste em crianças, constataram que mesmo as que tinham dificuldades de decodificação conseguiam acertar várias questões de compreensão. Então, resolveram testar o teste, ou seja, verificar até que ponto ele realmente avaliava a compreensão leitora. Para isso, administraram-no a alunos de graduação e a crianças de em média 11,3 anos. Os participantes receberam 65 perguntas de múltipla escolha sem os textos. Assim como no teste original, as perguntas foram introduzidas por uma frase que apresentava o tema do texto. Os resultados mostraram que 86% das questões foram respondidas com nível de acurácia acima do percentual de

adivinhação por estudantes de graduação. As crianças obtiveram desempenho inferior, porém ainda assim acima do percentual previsto de adivinhação. Os pesquisadores atribuem o resultado ao conhecimento prévio, concluindo que muitas questões podiam ser respondidas com base no conhecimento de mundo dos leitores, o que justifica o desempenho superior dos alunos de graduação. As questões chamadas “passage-independent” (independentes do texto) na verdade não avaliam a compreensão leitora, e sim o conhecimento de mundo dos leitores, conceito já explorado por autores como Van Dijk e Kintsch (1983). Os pesquisadores alertam para a necessidade de se fazer esse tipo de verificação em todos os testes de compreensão leitora, o que poderia ocorrer durante a fase de validação do teste ou por meio de um experimento piloto em que os participantes respondam as questões sem ler os textos.

O papel da memória também é um ponto importante de debate na avaliação da compreensão leitora, a começar pelas suas distintas funções. A memória de trabalho tem função operacional no gerenciamento das micro e macro proposições (Just & Carpenter, 1992), bem como nas operações de integração do conteúdo novo advindo do texto, lido ou ouvido, com o conteúdo relacionado, advindo da memória de longo prazo; assim que a integração ocorre, a memória de trabalho se libera para avançar a novos processos de integração. Na memória de curto prazo o leitor constrói a representação do modelo textual e situacional à medida que avança na leitura do texto. A memória de longo prazo atua com uma base de dados a ser acessada, por meio da qual o leitor completa as lacunas de sentidos no seu modelo textual; o conhecimento de mundo, incluindo o linguístico, armazenado nessa memória será o principal instrumento para os processos como previsão do conteúdo, integração dos sentidos globais e inferências.

É claro que não pode haver compreensão sem que haja a recuperação da base textual, contudo, o que mais além da lembrança precisa ser avaliado para se ter acesso à compreensão? Há diversos posicionamentos teóricos referentes a esse questionamento. Francis e colaboradores (2006), por exemplo, destacam três aspectos a serem examinados: o conhecimento, a aplicação e o engajamento, fazendo inclusive uma crítica aos métodos de avaliação, que na sua maioria avaliam apenas o primeiro aspecto. Outro parâmetro que pode ser adotado é o modelo de compreensão de Kintsch (1998, 2005), classificando as questões conforme o que acessam: base textual ou modelo situacional. A primeira consiste na representação do conteúdo proposicional, e a segunda nas relações construídas pelo leitor entre o a base textual e seu conhecimento de mundo, dependendo, assim, mais da memória de trabalho e da de longo prazo para a realização de processos de compreensão mais profundos como a integração e a inferência.

Há ainda o parâmetro proposto por Caldwell (2008), que comparou as categorias de questões adotadas por cinco pesquisadores, classificando-as em três grupos: questões literais, que seriam equivalentes ao conhecimento e à base textual, questões inferenciais textuais, que envolvem a inferência de conteúdos implícitos no texto, e questões de inferência aplicativa, que envolvem maior relação com o conhecimento de mundo, que originará o modelo situacional.

Esses três grupos assemelham-se ao modelo proposto por Francis et al. (2005), no entanto, nele a categoria de engajamento parece não ser contemplada, talvez devido à sua subjetividade, necessitando ser analisada com maior atenção. Uma classificação mais abrangente foi proposta por Marcuschi (1996). Partindo da concepção de leitura enquanto processo inferencial, ele sugeriu cinco horizontes de compreensão: falta de horizonte, horizonte mínimo, horizonte máximo, horizonte problemático e horizonte indevido. Eles são como níveis de compreensão que variam conforme o grau de interação do leitor com o texto.

O entendimento de que a compreensão leitora não se reduz apenas à lembrança da base textual é determinante para a fase de desenvolvimento da tarefa avaliativa. A começar pela permissão ou não de consulta ao texto durante a resolução das questões, pois, se não houver consulta haverá maior demanda de memória (Davey, 1990). Além disso, é importante que o pesquisador adote um modelo teórico para guiar a construção e a correção das tarefas. Por exemplo, se o avaliador decidisse construir uma tarefa com questões abertas, poderia adotar o modelo de Caldwell (2008) para definir quantas questões seriam literais, quantas inferenciais textuais e quantas seriam de inferência aplicativa. Já a proposta de Marcuschi (1996) parece mais adequada à análise de respostas a questões abertas, resumos e recontos do que para a definição do tipo de questão propriamente dito. As tarefas e a análise que se faz dos dados refletirão em alguma medida o posicionamento teórico de quem as constrói e suas concepções sobre a compreensão leitora.

O reconhecimento de palavras é um processo subjacente igualmente debatido. Ele é essencial para a compreensão, porém, modelos como Simple View of Reading (Hoover & Gough, 1990) defendem que, apesar de relacionados, reconhecimento de palavras e compreensão linguística são independentes. Keenan e colaboradores (2008) averiguaram a validade de quatro importantes testes de leitura padronizados e encontraram correlação moderada entre eles. A análise estatística mostrou que dois fatores independentes respondiam pela variação dos testes: a decodificação e a compreensão. Os testes, portanto, avaliavam diferentes componentes da habilidade leitora. Os autores apontaram como causa da diferença o formato das tarefas, sendo que o cloze e a identificação de figuras avaliavam mais o reconhecimento de palavras, enquanto múltipla escolha e questões abertas avaliavam mais a compreensão. Além disso, eles observaram que diferenças de desenvolvimento podem estar implicadas nessa variação, considerando que na fase inicial os leitores dependem mais dos processos de decodificação e, à medida que os automatizam, a compreensão ganha maior autonomia. Tais observações apontam para a importância de se construírem métodos de avaliação mais sensíveis às diferenças desenvolvimentais da leitura.

Adlof e colaboradores (2011) também notaram a diferença na sensibilidade dos testes de leitura quanto às habilidades componentes, além de destacarem a natureza do próprio texto que, nas séries iniciais, normalmente é narrativo e possui linguagem mais semelhante à oral, demandando baixa compreensão. Eles sugerem que se desenvolvam testes de leitura capazes de detectar problemas de compreensão já

nas fases iniciais de alfabetização, de modo que se possam iniciar os processos de intervenção o mais cedo possível. Pesquisadores como Cain e Oakhill (2006) recomendam que os leitores sejam pareados conforme a habilidade de reconhecimento de palavras por meio de teste específico, de modo a garantir que se esteja avaliando predominantemente a compreensão. Uma das grandes falhas na avaliação do desempenho leitor nas séries iniciais é não se avaliarem a decodificação e a compreensão separadamente. Na maioria das vezes identificam-se leitores com dificuldades de decodificação, mas não leitores com dificuldades específicas de compreensão, que só serão notados nas séries finais, retardando, assim, as intervenções.

Já existem tentativas de se avaliar os subprocessos da leitura separadamente. Sabatini e colegas (2013) construíram para isso seis subtestes: reconhecimento de palavra e decodificação, vocabulário, consciência morfológica, processamento de sentenças, compreensão básica e compreensão avançada. Eles afirmam que a aplicação de subtestes é uma forma eficaz de identificar onde os leitores estão falhando e assim construir formas de intervenção mais eficientes. Cain e Oakhill (2012) reconhecem a importância de acessar separadamente as habilidades componentes da compreensão leitora, entretanto, afirmam que essa é uma tarefa bastante complexa devido a forte interrelação entre as habilidades.

Como vimos, todas as tarefas possuem suas limitações. A validade e a confiabilidade estão entre os aspectos de maior preocupação entre os pesquisadores, pois objetiva-se construir instrumentos eficazes que, ao mesmo tempo, tenham alguma validade ecológica e validade de construto. Por isso, ao optar por uma tarefa de avaliação da compreensão leitora é preciso primeiramente considerar a concepção de leitura adotada no estudo e o objetivo da tarefa: diagnóstico, verificação da eficiência de métodos de instrução, análise do desenvolvimento da compreensão leitora, avaliação do grau de compreensão, comparação de níveis de proficiência leitora entre grupos, etc. Em seguida é importante analisar as limitações de cada tipo de tarefa, as demandas cognitivas de cada uma delas, os critérios para a correção das respostas e os parâmetros teóricos para a interpretação dos resultados.

Além dos processos cognitivos subjacentes aqui mencionados, poderíamos destacar ainda a atenção, a capacidade de raciocínio, a experiência de leitura e a motivação. Pesquisas têm demonstrado que a baixa atenção e o déficit na memória de trabalho são fatores originadores de problemas de compreensão (McVay & Kane, 2012). Da mesma forma, Logan, Medford e Hughes (2011) obtiveram evidências de que o nível de motivação intrínseca contribui de forma significativa para o desempenho na compreensão textual de alunos com baixa performance leitora. Apesar das evidências, ainda é preciso avançar muito na pesquisa de fatores individuais como motivação e interesse do leitor. Como veremos na seção seguinte, determinar a leiturabilidade dos textos utilizados na construção de tarefas de avaliação da compreensão leitora requer que se considerem além de fatores linguísticos presentes no texto, outros fatores cognitivos e subjetivos como competência e motivação do leitor.



*Leiturabilidade dos textos*

Outro desafio que encontramos ao construir tarefas de avaliação da compreensão leitora é selecionar textos com leiturabilidade adequada para o grupo de leitores que se quer avaliar. A leiturabilidade, também designada na literatura como inteligibilidade e compreensibilidade, consiste no nível de leitura exigido para a compreensão do texto, ou em outras palavras, o nível de dificuldade textual (Collins-Thompson, 2014). Há pelo menos 90 anos busca-se identificar os melhores critérios para determinar a leiturabilidade dos textos, sejam eles critérios quantitativos como o cálculo do número total de palavras, número de palavras por sentenças, número de sentenças, frequência das palavras do texto, número de palavras de conteúdo e de classes fechadas, grau de complexidade sintática, etc.; ou critérios qualitativos que dependem em grande parte da análise de avaliadores sobre aspectos linguísticos mais complexos como coesão, coerência, clareza, concisão, relevância, entre outros.

Os primeiros estudos que buscaram avaliar a leiturabilidade datam de 1920 nos Estados Unidos (Dubay, 2004). Por meio de métodos estatísticos esses estudos desenvolveram fórmulas para pontuar a leiturabilidade dos textos. Esses métodos, alguns dos quais ainda hoje utilizados, possuem alta validade e classificam os textos em sua equivalência aos anos escolares, considerando, para isso, critérios como o número e a extensão de palavras e sentenças. Os mais empregados são os propostos por Flesch (1949), Chall (1958), Chall e Dale (1995), Coleman (1965), bem como os métodos Forcast (Caylor, Sticht, Fox & Ford, 1973), Smog (McLaughlin, 1969), Flesch-Kincaid (Kincaid, Fishburne, Rogers & Chissom, 1975) e Atos (School Renaissance Institute, 2000).

Ao revisar as novas tendências no cálculo da leiturabilidade, Collins-Thompson (2014) identificou nos métodos atuais a incorporação de um número muito maior de métricas linguísticas como frequência de palavra, números de pronomes e preposições, densidade proposicional, e ainda de recursos computacionais como: sistemas orientados por dados (data driven), sistemas baseados em conhecimento (knowledge-based), aprendizado de máquina (machine learning), corpus linguísticos, que permitiram uma maior sofisticação e eficiência na análise automática da leiturabilidade. Lexile, por exemplo, é uma das ferramentas mais atuais na língua inglesa disponível na internet para pesquisadores, professores e até mesmo pais. Ele considera a frequência das palavras e a extensão das frases (Harvey, 2011), o que garante um resultado ainda mais eficiente. No site (<http://www.lexile.com/>) é possível buscar livros adequados a diversos níveis de compreensão classificados do primeiro ao décimo segundo ano escolar.

Infelizmente, não há nenhuma fórmula para o cálculo da leiturabilidade desenvolvida exclusivamente para a língua portuguesa brasileira. Encontramos apenas dois estudos que reportamos brevemente. No primeiro, Martins e colaboradores (1996) avaliaram a aplicabilidade da fórmula de Flesch (1949) a textos em português retirados de livros didáticos das áreas sociais e científicas do ensino fundamental, médio e universitário. Ao todo foram analisados 10 textos de em média 150 palavras para cada ano escolar. Eles descobriram

que, devido à diferença na extensão das palavras nas duas línguas, a fórmula apresentava valores inferiores quando aplicada aos textos em português brasileiro, por isso, adicionaram 42 pontos ao escore final e obtiveram bons resultados. No segundo estudo, Martins e Filgueiras (2007) buscaram avaliar a apreensibilidade da informação textual em sites do governo através da fórmula de Flesch-Kincaid, uma versão atualizada e simplificada da fórmula original desenvolvida por Kincaid et al. (1975). Ao todo avaliaram 24 textos com no mínimo 250 palavras. Eles também utilizaram a técnica cloze para avaliar a leiturabilidade de dois textos da amostra. Cada texto foi lido e completado por cinco leitores. As duas avaliações mostraram que o nível de leiturabilidade dos textos era baixo sendo, portanto, considerados difíceis quando comparados à média de leitura da população brasileira. Os dois textos avaliados por meio do cloze foram simplificados e reavaliados por meio da mesma técnica, apresentando grau de leiturabilidade mais alto após a sua adaptação.

Em 2007, um grupo de pesquisadores liderados por Sandra Aluísio (USP) iniciou o projeto PorSimples (Simplificação Textual do Português para Inclusão e Acessibilidade Digital) cujo fim é tornar a informação mais acessível a analfabetos funcionais, leitores disléxicos ou com dificuldades de compreensão e aprendizagem, para isso, construíram um programa capaz de simplificar textos. Primeiramente, ele analisa a inteligibilidade dos textos (Scarton & Aluísio, 2010) por meio de métricas selecionadas do Coh-Matrix-Port, e então aplica métodos de simplificação sintática, encurtamento de sentenças, e substituição lexical. O recurso está disponível gratuitamente na internet (<http://www.nilc.icmc.usp.br/porsimples/simplifica/>), podendo ser facilmente utilizado por pesquisadores, escritores, professores e público em geral. Ele fornece três níveis de classificação dos textos de acordo com o Indicador de Alfabetismo Funcional INAF/Brasil – 2007, considerando os níveis de letramento rudimentar, básico e pleno. Outras ferramentas interessantes foram desenvolvidas no projeto como o Facilita que produz uma versão simplificada dos textos de forma a facilitar a compreensão dos leitores, e o Simplifica cuja finalidade é ajudar autores na escrita de textos mais simples.

O Coh-Matrix-Port (<http://www.nilc.icmc.usp.br:3000/>) é uma ferramenta online também desenvolvida no projeto PorSimples, derivado da versão original em inglês. Por meio dele é possível obter a análise da leiturabilidade através da fórmula de Flesch corrigida para o português (Aluísio, Specia, Gasperin & Scarton, 2010). O resultado da análise indica o nível de dificuldade do texto (muito fácil, fácil, difícil, muito difícil) e sua correspondência ao nível de escolaridade (até quarta série do ensino fundamental, até oitava série do ensino fundamental, ensino médio ou universitário, áreas acadêmicas específicas). Além disso, a ferramenta possui 60 métricas de análise linguística textual: lexical, sintática, discursiva e conceitual. Rodrigues, Freitas e Quental (2013) explicam como utilizaram as métricas do Coh-Matrix-Port, inclusive o índice de leiturabilidade de Flesch, para analisar a adequação de livros literários distribuídos para jovens e adultos recém-alfabetizados.

O Coh-Metrix pode ser uma importante ferramenta na análise de textos, escritos ou orais, produzidos não só por participantes saudáveis, foco deste estudo, como também por populações clínicas, como afásicos, pacientes com doenças psiquiátricas ou neurodegenerativas. Suas métricas permitem a identificação de padrões linguísticos e características discursivas específicas de um determinado grupo de participantes que podem ser comparados entre si ou com grupos controle.

As tradicionais fórmulas de leiturabilidade, citadas anteriormente, apesar de obterem alta validade, não avaliam o nível de coerência dos textos. McNamara, Graesser e Louwerse (2012) afirmam que a avaliação dos níveis lexicais e sentenciais é insuficiente para prever o nível de dificuldade do texto, pois, mesmo que utilizássemos uma sequência de palavras desconexas, poder-se-ia calcular a leiturabilidade, apesar da ausência de sentido. Os pesquisadores examinaram diferentes tipologias textuais: texto narrativo, expositivo da área das humanas e expositivo da área de ciências naturais usados em vários níveis escolares; analisando fatores tradicionais de leiturabilidade e fatores de coerência: coesão referencial local e global, coesão verbal, uso de conectivos e coesão causal por meio do Coh-Metrix americano. Eles concluíram que a facilidade dos textos narrativos provém do uso de palavras mais frequentes e de conectivos, embora sejam menos coesos. Já os textos expositivos de ciências naturais são desafiadores porque contêm palavras menos frequentes e menor número de conectivos, ainda que tentem

compensar com maior referenciação e coesão verbal. Os textos expositivos da área das humanas surpreenderam ao apresentarem as dificuldades encontradas nas outras duas tipologias: complexidade conceitual, longas frases, número moderado de conectivos e pouca coesão.

A análise de fatores de coerência, como vemos, é enriquecedora ao fornecer informações sobre características linguísticas que variam em função dos tipos textuais (como o narrativo e o argumentativo), dos gêneros textuais (incluindo, por exemplo, contos, crônicas, textos de divulgação de conteúdo científico) e do ano escolar, devendo, assim, ser consideradas na avaliação da leiturabilidade. Além disso, os autores apontam para variações na forma como esses fatores afetam diferentes leitores. Os mais habilidosos conseguem muitas vezes superar as lacunas no conteúdo textual por meio de seu conhecimento prévio, sendo, portanto, menos dependentes desses fatores. Já os leitores menos habilidosos são mais dependentes da explicitude das relações micro e macrotextuais.

Além dos fatores linguísticos que influenciam diretamente a leiturabilidade, Klare (1976) chama a atenção para as diferenças individuais dos leitores como conhecimento prévio, proficiência em leitura, motivação, maturidade e nível intelectual, que devem ser avaliados em comparação com o conteúdo do texto. Na Tabela 2, resumimos alguns dos fatores debatidos por Klare (1976), juntamente com outros aqui citados, a serem considerados na avaliação da leiturabilidade de textos.

Tabela 2. Fatores para a análise da leiturabilidade considerando o texto e o leitor

Leiturabilidade textual					
Palavras	Sentenças	Conteúdo	Discurso	Competência ao leitor	Motivação ao leitor
Proporção de palavras de conteúdo X função	Extensão	Profundidade do conteúdo	Gênero textual	Conhecimento do tema	Motivação intrínseca
Qualidade do conteúdo das palavras: frequência, fototipicidade, polifonia*, familiaridade, extensão e termos técnicos	Ativas X passivas	Densidade de novas informações e argumentos	Coerência	Proficiência em leitura: uso de estratégias cognitivas/metacognitivas	Motivação extrínseca
Concretude X abstração	Afirmativas X negativas	Densidade proposicional	Coesão	Nível intelectual e capacidade cognitiva como atenção, memória de trabalho e monitoramento	Engajamento com o texto
Valor associativo	Encaixadas X não-encaixadas	Explicitude de ideias	Estilo: uso de figuras de linguagem	Nível de maturidade e experiência leitora	Tempo disponível e usado
Construções verbais ativas X nominalizadas	Baixa profundidade X alta profundidade	Complexidade pragmática	Fonte e destinatário	Familiaridade com o gênero textual	Preferências e interesses

*Nota.* \* Palavras de conteúdo são em geral substantivos, verbos, adjetivos, advérbios e pronomes interrogativos. Palavras funcionais são artigos, preposições, conjunções e pronomes pessoais. Palavras prototípicas são as que melhor representam um grupo semântico, ex: fruta – maçã. Palavras polissêmicas são as que possuem múltiplos significados, ex: direito (área do conhecimento, direção, certo, etc).



Com o objetivo de verificar o impacto da leitura sobre a compreensão, Klare (1976) investigou o efeito que mudanças nos textos têm sobre a compreensão dos leitores, identificando fatores que interagem com a leitura, de forma que leitores com grande conhecimento prévio e/ou motivação e/ou inteligência facilmente superam as dificuldades de textos com baixo nível de leitura. Mostrou, assim, que a leitura não tem um valor único para todos os leitores e, por isso, seu cálculo deve considerar não apenas o texto em si mesmo, mas o texto em relação aos leitores, ou seja, em relação às especificidades do grupo de leitores que será avaliado. O pesquisador propõe que o desempenho do leitor varie em função da situação do teste; essa, por sua vez, interage com a competência e a motivação, das quais depende a avaliação do conteúdo e do nível de leitura. Apesar de a motivação interferir na leitura, ela não pode ser diretamente controlada pelo pesquisador porque é subjetiva ao leitor. Todavia, modificações na tarefa como extensão do texto, formas de instruções, apresentação (papel x computador) e escolha dos temas, ou mesmo na situação do teste como duração, horário e local, podem contribuir para diminuir a tensão e motivar os participantes para seu engajamento na realização da tarefa.

Outra proposta de avaliação do nível de dificuldade de textos foi feita por Irwin e Davis (1980), que elaboraram uma lista de perguntas para nortear a avaliação de livros didáticos (Readability Checklist), partindo de dois critérios principais: a compreensibilidade e a aprendibilidade. Os livros são pontuados numa escala de 1 a 5 em cada pergunta, que pode ser adaptada para a avaliação de textos isolados. Análises qualitativas como essa podem ser adotadas se previamente estabelecidos critérios claros, podendo esse ser um método confiável quando adotado o processo de avaliação cega por pares em que no mínimo dois avaliadores pontuam, depois comparam suas avaliações e buscam chegar a um consenso.

Determinar o nível de leitura de um texto é uma árdua tarefa na construção de avaliações da compreensão leitora devido às múltiplas dimensões a serem consideradas. Os primeiros métodos levavam em conta apenas fatores linguísticos quantificáveis; contudo, fatores concernentes à profundidade do conteúdo, densidade de novas informações, densidade de argumentos, concretude e imageabilidade das palavras, explicitude das ideias (Johnston, 1983), número de proposições (Kintsch, 1998), gênero textual, coesão e coerência (Mcnamara et al., 2012), intertextualidade (Gabriel, 2005), aspectos pragmáticos (Flores, 2007), diferenças culturais e individuais (Marcuschi, 1985; Kleiman, 1998, 2002), dentre outros aspectos, tornam a tarefa mais complexa e menos objetiva. A combinação de métodos de quantificação com métodos qualitativos como o de julgamento parece gerar uma avaliação mais ampla e confiável. Conforme Campbell (2005), as variações entre os textos e as tarefas são tão grandes que utilizar um único trecho ou uma única forma de resposta não é confiável. Por isso, muitos testes padronizados, ao invés de usarem apenas um texto longo, preferem diversificar os gêneros e usar textos mais curtos, tentando assim minimizar efeitos advindos, por exemplo, do

conhecimento prévio, motivação do leitor e características próprias dos gêneros textuais.

### Considerações Finais

Neste artigo exploramos dois importantes fatores a serem considerados na avaliação da compreensão leitora: demandas cognitivas subjacentes às tarefas avaliativas e leitura textual. Os desafios na construção de tarefas de avaliação da compreensão leitora são inúmeros. Tratamos aqui de apenas dois, pois consideramos serem pouco debatidos especialmente na literatura em língua portuguesa.

A escolha do texto apropriado é uma fonte de grandes variações (Tabela 2) que, no entanto, pode ser mais bem controlada, comparada às variações cognitivas individuais, graças às tecnologias desenvolvidas pela linguística de corpus e linguística computacional, que têm evoluído no Brasil nas últimas décadas. O projeto PorSimples, que citamos anteriormente, é um exemplo louvável de como a tecnologia pode auxiliar na avaliação da leitura, permitindo uma escolha mais cuidadosa de textos, não só para a área da pesquisa em leitura, mas também para a editoração de livros didáticos, revistas, jornais, manuais de instrução, legislação, provas de avaliação escolar e tantos outros domínios que fazem uso do texto escrito, podendo alcançar um grande impacto social ao tornar a informação mais acessível a leitores com dificuldades. A temática dos textos que compõe as tarefas, como vimos anteriormente, deve ser cuidadosamente escolhida juntamente com as questões de compreensão leitora, de modo que se evitem elaborar perguntas que possam ser respondidas sem a leitura do texto, com base apenas no conhecimento prévio do leitor acerca do tema.

Quanto ao formato da tarefa, verificamos que não há consenso sobre qual seja a melhor forma, visto que a escolha depende, antes de tudo, da concepção de leitura que se adota e dos objetivos da avaliação. Todas as tarefas possuem limitações, cabendo ao pesquisador decidir qual será a mais informativa de acordo com os objetivos da avaliação. Escolher a mais apropriada requer que se analise primeiramente a finalidade, por que queremos avaliar a compreensão, o número de participantes, o tempo disponível, os recursos e os investimentos necessários. O ideal é combinar mais de um método e obter diversas amostras do desempenho da compreensão leitora de cada participante.

Outro desafio é a confiabilidade e a validade dos testes o que, muitas vezes, impede o uso de métodos alternativos de avaliação (Valencia & Pearson, 1987; Pearson & Hamm, 2005) quando a finalidade é a pesquisa científica. Os esforços para ganhar confiabilidade normalmente sacrificam a validade dos testes (Johnston, 1983). Dessa forma, a busca por um resultado mais objetivo e confiável, menos suscetível a erros, acaba gerando tarefas artificiais que não reproduzem a situação e os processos de leitura cotidianos. Em vista disso, Graesser e Hu (2012) sugerem aspectos a serem melhorados nas tarefas de avaliação da compreensão leitora que consideramos relevantes de serem aqui citados: avaliação ao longo de um período, maior diversidade de textos, integração do texto com recursos multimídia, diversificação de tarefas, verificação dos

processos de leitura, medidas que alcancem níveis de compreensão mais profundos, medidas que contemplem diversos construtos de compreensão leitora, métodos psicométricos que incorporem medidas e parâmetros de prospecção de dados (data mining) e avaliações informativas que forneçam o perfil detalhado dos leitores. Algumas alterações poderiam ser facilmente implementadas, como a utilização de textos de gêneros diversos e de maior controle do nível de leitorabilidade dos textos. Outras poderiam se beneficiar com o uso de recursos tecnológicos como a obtenção de medidas mais profundas de compreensão e verificação dos processos de compreensão leitora.

Embora a pesquisa sobre leitura tenha feito importante progresso nas últimas décadas, as tarefas de avaliação da compreensão leitora têm se modificado lentamente e incorporado poucos avanços. Conforme Mislevy e Sabatini (2012) argumentam, precisamos construir formas de acesso à compreensão com base em uma visão mais atual de leitura considerando tanto os aspectos sociointeracionais quanto cognitivos envolvidos na compreensão leitora. Van den Broek e Espin (2012), por exemplo, elaboraram uma breve fundamentação teórica que se propõe a guiar o desenvolvimento e a atualização de tarefas de avaliação da compreensão leitora tomando como referência o modelo integrado de compreensão leitora. Dentre suas sugestões, destaca-se a incorporação de tarefas que verifiquem a sensibilidade à centralidade estrutural, ou seja, em que medida o leitor é capaz de identificar e representar informações centrais no texto.

Francis e colegas (2005, p. 376) afirmam que a imperfeição é inerente às formas de acesso à compreensão e que suas diferenças refletem nada mais do que variações no grau de imperfeição. Isso não quer dizer que devamos desistir de avaliar a compreensão leitora, mas que precisamos de mais pesquisa e investimento para aperfeiçoar as formas de avaliação. Ademais, é urgente que se revitalize essa área de estudo, lembrando que todas as investigações envolvendo a leitura estão condicionadas à sua forma de avaliação. No Brasil a situação é ainda mais crítica porque temos pouca pesquisa sobre a avaliação da compreensão leitora. O desenvolvimento dessa área de estudo é fundamental para que a pesquisa em leitura avance em nosso país, sendo para isso necessário o trabalho colaborativo de profissionais das diversas áreas do conhecimento que estudam a leitura e a escrita. Os baixos índices de desempenho em leitura em grande parte dos países da América Latina nos convocam a, como neste artigo, refletir sobre os instrumentos existentes e sobre as formas de aperfeiçoá-los, tendo em vista a construção de tarefas avaliativas que possam auxiliar tanto na pesquisa sobre a compreensão leitora quanto no diagnóstico das dificuldades de compreensão em leitura visando a práticas de intervenção educacional e clínica.

### Referências

Aluísio, S., Specia, S., Gasperin, C., & Scarton, C. (2010). Readability assessment for text simplification. *Proceedings of the Fifth Workshop on Innovative*

*Use of NLP for Building Educational Applications - NAACL HLT*, 1-9, Los Angeles, CA, USA.

- Adlof, S. M., Perfetti, C. A., & Catts, H. W. (2011). Developmental changes in reading comprehension: implications for assessment and instruction. In S. J. Samuels, & A. E. Farstrup (Eds.), *What research has to say about reading instruction* (pp. 186-214). International Reading Association: Newark, DE.
- Bennett, R. E., & Ward, W. C. (1993). *Construction versus choice in cognitive measurement: issues in constructed response, performance testing, and portfolio assessment*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Berk, R. (1996). A consumer's guide to multiple choice item formats that measure complex cognitive outcomes. In National evaluation systems (Eds.), *From policy to practice* (pp. 101-127). Amherst, MA: Pearson Publishing.
- Brooks, M., Basu, S., Jacobs, C., & Vanderwende, L. (2014). Divide and correct: using clusters to grade short answers at scale. *Proceedings of the 1st Conference on Learning at Scale*, 89 – 98, Atlanta, GA, USA.
- Cain, K., & Oakhill, J. (2006). Assessment matters? Issues in the measure of reading comprehension. *British Journal of Educational Psychology*, 76(4), 697–708. doi: 10.1348/000709905X69807.
- Cain, K., & Oakhill, J. (2012). Reading comprehension development from seven to fourteen years: implications for assessment. In J. P. Sabatini, E. Albro, & T. O'reilly (Eds.), *Measuring up: advances in how we assess reading ability* (pp. 697–708). Plymouth: Rowman & Littlefield Publishers.
- Caldwell, J. S. (2008). *Comprehension assessment: a classroom guide*. The Guilford Press: New York.
- Campbell, J. R. (1999). *Cognitive processes elicited by multiple-choice and constructed-response questions on an assessment of reading comprehension*. Tese não publicada, Temple University, Philadelphia, United States.
- Campbell, J. R. (2005). Single instrument, multiple measures: considering the use of multiple item formats to assess reading comprehension. In S. G. Paris, & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 347-368). Lawrence Erlbaum Associates: New Jersey.
- Carlisle, J. (1989). Diagnosing comprehension deficits through listening and reading. *Annals of Dyslexia*, 39(1), 159-176. doi: 10.1007/BF02656907

- Caylor, J. S., Sticht, T. G., Fox, L. C., & Ford, J. P. (1973). *Methodologies for determining reading requirements of military occupational specialties* (technical report n. 73-5). Human Resources Research Organization: Alexandria, VA.
- Chall, J. S. (1958). *Readability: an appraisal of research and application*. Ohio State University Press: Columbus, OH.
- Chall, J. S., & Dale, E. (1995). *Readability revisited, the new Dale-Chall readability formula*. Brookline Books: Cambridge, MA.
- Choppin, B. F., & Purves, A. L. (1969). A comparison of open-ended and multiple-choice items dealing with literacy understanding. *Research in the Teaching of English*, 3(1), 15-24.
- Coleman, E. B. (1965). *On understanding prose: some determiners of its complexity*. (NSF Final Report GB-2604). National Science Foundation: Washington, D.C.
- Collins-Thompson, K. (2014). Computational assessment of text readability: a survey of current and future research. *International Journal of Applied Linguistics* 165(2), 97-135. doi: [10.1075/itl.165.2](https://doi.org/10.1075/itl.165.2)
- Davey, B. (1990). Field dependence-independence and reading comprehension questions: task and reader interactions. *Contemporary Educational Psychology*, 15(3), 241-250. doi: [10.1016/0361-476X\(90\)90021-R](https://doi.org/10.1016/0361-476X(90)90021-R)
- Dubay, W. H. (2004). *The Principles of Readability*. Costa Mesa, CA: Impact Information. Recuperado em 06 março, 2015, de <http://www.impact-information.com/impactinfo/readability02.pdf>
- Farr, R., & Carrey, R. F. (1986). *Reading: what can be measured?* International Reading Association: Newark.
- Flesch, R. (1949). *The art of readable writing*. Harper: New York.
- Fletcher, J. M. (2006). Measuring reading comprehension. *Scientific Studies of Reading*, 10(3), 323-330. doi: [10.1207/s1532799xssr1003\\_7](https://doi.org/10.1207/s1532799xssr1003_7)
- Flores, O. C. (2007). Como avaliar a compreensão leitora? *Signo*, 32(53), 54-65. doi: <http://dx.doi.org/10.17058/signo.v32i53.123>
- Fonseca, R. P., Casarin, F. S., Oliveira, C. R., Gindri, G., Soares, E. C. S., Ortiz, K. Z., Parente, M. A. M. P., & Scherer, L. C. (2011). Adaptação de instrumentos neuropsicológicos verbais: um fluxograma de procedimentos para além da tradução. *Interação em Psicologia*, 15(n.especial), 59-69.
- Francis, D. J., Fletcher, J. M., Catts, H. W., & Toblin, J. B. (2005). Dimensions affecting the assessment of reading comprehension. In S. G. Paris, & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 369-394). Lawrence Erlbaum Associates: New Jersey.
- Francis, D. J., Snow, C. E., August, D., Carlson, C. D., Miller, J., & IGLESIAS, A. (2006). Measures of reading comprehension: a latent variable analysis of the diagnostic assessment of reading comprehension. *Scientific Studies of Reading*, 10(3), 301-322. doi: [10.1207/s1532799xssr1003\\_6](https://doi.org/10.1207/s1532799xssr1003_6)
- Gabriel, R. (2005). Compreensão em leitura: como avaliá-la. In A. Olmi, & N. Perkoski (Eds.), *Leitura e Cognição: uma abordagem transdisciplinar* (pp. 165-210). Edunisc: Santa Cruz do Sul.
- Graesser, A., & Hu, X. (2012). Moving forward on reading assessment. In J. P. Sabatini, E. Albro, & T. O'reilly (Eds.), *Measuring up: advances in how we assess reading ability* (pp. 153-157). Plymouth: Rowman & Littlefield Publishers.
- Harvey, C. A. (2011). An inside view of Lexile measures: an interview with Malbert Smith III. *Knowledge Quest*, 39(4), 56-59.
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: an interdisciplinary journal*, 2(2), 127-160. doi: [10.1177/0022219411432685](https://doi.org/10.1177/0022219411432685)
- Irwin, J. W., & Davis, C. A. (1980). Assessing Readability: the checklist approach. *Journal of Reading*, 24(2), 124-130.
- Johnston, P. H. (1983). *Reading comprehension assessment: a cognitive basis*. International Reading Association: Newark.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review*, 99(1), 122-149.
- Keenan, J. M., & Betjemann, R. S. (2006). Comprehending the Gray Oral Reading Test without reading it: why comprehension tests should not include passage-independent items. *Scientific Studies of Reading*, 10(4), 63-380. doi: [10.1207/s1532799xssr1004\\_2](https://doi.org/10.1207/s1532799xssr1004_2)
- Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12(3), 281-300. doi: [10.1080/10888430802132279](https://doi.org/10.1080/10888430802132279)
- Kincaid, J. P., Fishburne, R. P., Rogers, R., & Chissom, B. S. (1975). *Derivation of new readability formulas*

- (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel. Research Branch Report.
- Kintsch, W. (1998). *Comprehension: a paradigm for cognition*. Cambridge University Press: Cambridge.
- Kintsch, E. (2005). Comprehension theory as a guide for the design of thoughtful questions. *Top Lang Disorders*, 25(1), 51-64. doi: [10.1097/00011363-200501000-00006](https://doi.org/10.1097/00011363-200501000-00006)
- Klare, G. R. (1976). A second look at the validity of readability formulas. *Journal of Literacy Research*, 8(2), 129-152. doi: [10.1080/10862967609547171](https://doi.org/10.1080/10862967609547171)
- Kleiman, A. (1998). *Oficina de leitura: teoria e prática* (6ª ed). Pontes: Campinas.
- Kleiman, A. (2002). *Texto e leitor: aspectos cognitivos da leitura* (8ª ed). Pontes: Campinas.
- Kobayashi, M. (2009). *Hitting the mark: how can text organization and response format affect reading test performance?* Bern: Peter Lang.
- Langer, J. A. (1987). The construction of meaning and the assessment of comprehension: an analysis of reader performance on standardized test items. In R. Freedle (Ed.), *Cognitive and linguistic analyses of standardized test performance* (pp. 2-38). Norwood, New Jersey: Ablex.
- Logan, S., Medford, E., & Hughes, M. (2011). The importance of intrinsic motivation for high and low ability readers' reading comprehension performance. *Learning and Individual Differences*, 21(1), 124-128. doi: [10.1016/j.lindif.2010.09.011](https://doi.org/10.1016/j.lindif.2010.09.011)
- Marcuschi, L. A. (1985). Leitura como processo inferencial num universo cultural cognitivo. *Leitura, Teoria e Prática*, 4(5), 1-14.
- Marcuschi, L. A. (1996). Exercícios de compreensão ou cópiação nos manuais de ensino de língua? *Em Aberto*, 16(69), 64-82.
- Martins, S., & Filgueiras, F. (2007). Métodos de avaliação de apreensibilidade das informações textuais: uma aplicação em sítios de governo eletrônico. *Proceedings of the Workshop on Perspectives, Challenges and Opportunities for Human-Computer Interaction in Latin America*, Rio de Janeiro, RJ, Brasil.
- Martins, T. B. F., Ghiraldelo, C. M., Nunes, M. G. V., & Oliveira, O. N. J. (1996). Readability formulas applied to textbooks in Brazilian Portuguese. *Notas do ICMSC*. São Carlos: USP.
- Mclaughlin, G. H. (1969). SMOG grading - a new readability formula. *Journal of reading*, 22(8), 639-646.
- Mcnamara, D. S., Graesser, A. C., & Louwerse, M. M. (2012). Sources of text difficulty: across the ages and genres. In J. P. Sabatini, & E. Albro (Eds.), *Assessing reading in the 21st century: aligning and applying advances in the reading and measurement sciences*. Lanham, MD: R&L Education.
- Mcvay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *The Journal of Experimental Psychology*, 141(2), 302-320. doi: [10.1037/a0025250](https://doi.org/10.1037/a0025250)
- Mislevy, R. L., & Sabatini, J. P. (2012). How research on reading and research on assessment are transforming reading assessment (or if they aren't how they ought to). In J. P. Sabatini, E. Albro, & T. O'reilly (Eds.), *Measuring up: advances in how we assess reading ability* (pp. 153-157). Plymouth: Rowman & Littlefield Publishers.
- Nation, K., & Snowling, M. (1997). Assessing reading difficulties: the validity and utility of current measures of reading skill. *British Journal of Educational Psychology*, 67(3), 359-370. doi: [10.1111/j.2044-8279.1997.tb01250.x](https://doi.org/10.1111/j.2044-8279.1997.tb01250.x)
- Nunnally, J. C. (1959). *Tests and measurements: assessment and prediction*. New York: McGraw-Hill Book Company.
- Oakhill, J., Cain, K., & Elbro, C. (2014) *Understanding and teaching reading comprehension: a handbook*. Routledge, London.
- Ozuru, Y., Briner, S., Kurby, C. A., & McNamara, D. S. (2013). Comparing comprehension measured by multiple-choice and open-ended questions. *Canadian Journal of Experimental Psychology*, 67(3), 215-227. doi: [10.1037/a0032918](https://doi.org/10.1037/a0032918)
- Pearson, P. D., & Hamm, D. N. (2005). The assessment of reading comprehension: a review of practices – past, present and future. In S. G. Paris, & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 13-70). New Jersey: Lawrence Erlbaum Associates.
- Priebe, S. J., Keenan, J. M., & Miller, A. C. (2011). How prior knowledge affects word identification and comprehension. *Reading and writing*, 25(1), 581-586. doi: [10.1007/s11145-010-9260-0](https://doi.org/10.1007/s11145-010-9260-0)
- Rodrigues, E., Freitas, C., & Quental, V. (2013). Análise de inteligibilidade textual por meio de ferramentas de processamento automático do português: avaliação da Coleção Literatura para Todos. *Letras de Hoje*, 48(1), 91-99.

- Sabatini, J. P., Bruce, K., & Steinberg, J. (2013). *SARA reading components tests, RISE form: test design and technical adequacy*. ETS Research Report Series. doi: 10.1002/j.2333-8504.2013.tb02315.x
- Scarton, C. E., & Aluísio, S. M. (2010). Análise da inteligibilidade de textos via ferramentas de processamento de língua natural: adaptando as métricas do Coh-Metrix para o português. *Linguamática*, 2(1), 45-62.
- School Renaissance Institute. (2000). *The ATOS readability formula for books and how it compares to other formulas*. Madison, WI: School Renaissance Institute, Inc.
- Sweet, A. P. (2005). Assessment of reading comprehension: the RAND Reading Study Group Vision. In S. G. Paris, & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 3-12). New Jersey: Lawrence Erlbaum Associates.
- Valencia, S. W., & Pearson, P. O. (1987). Reading assessment: time for a change. *The Reading Teacher*, 40(8), 726-732.
- van den Broek, P., & Espin, C. A. (2012). Connecting cognitive theory and assessment measuring individual differences in reading comprehension. *School Psychology Review*, 41(3), 315-325.
- van den Bergh. (1990). On the construct validity of multiple choice items for reading comprehension. *Applied Psychological Measurement*, 14(1), 1-12. doi. 10.1177/014662169001400101
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Wagner, G. (1986). Interpreting cloze scores in the assessment of text readability and reading comprehension. *Directions: Journal of Educational Studies*, 8(16), 68-72.
- Wiederholt, J. L., & Bryant, B. R. (2001). *GORT 4: Gray Oral Reading Tests Examiner's manual*. Austin, TX: PRO-ED.