

POLICY REVIEW: Medindo competência de leitura

Autores:



Guilherme Lichand
*Professor de Educação
na Stanford GSE*



Maria E. Perpétuo
*Assistente de Pesquisa
no CCWD*



Mavigson Silva
*Assistente de Pesquisa
no CCWD*



Lucas Klotz
*Assistente de Pesquisa
no CCWD*

O desafio brasileiro

O Brasil está finalmente definindo o que significa ser alfabetizado ao final do 2º ano do Ensino Fundamental. Se a Base Nacional Curricular Comum (BNCC) define o conjunto de habilidades e competências que uma criança deve aprender a cada série, não explicita quais uma criança deveria dominar para ser considerada proficiente. De fato, isso não é tarefa simples: envolve definir padrões mínimos para diferentes habilidades – como o número de palavras lidas corretamente por minuto, uma medida de fluência leitora, ou os tipos de texto que um aluno deveria ser capaz de ler ou escrever. Como exemplo, Portugal define a fluência adequada nessa faixa etária como de ao menos 75 palavras por minuto. No Brasil, o Ministério da Educação (MEC) recém anunciou que, ao final do 2º ano, ser alfabetizado deveria corresponder a “escrever bilhetes e convites e ler textos simples, tirinhas e histórias em quadrinhos” – correspondente a uma pontuação de ao menos 743 pontos na avaliação do Sistema Nacional de Avaliação Básica (Saeb).

Tanto no caso da fluência quanto da proficiência, o desafio é que esses números não fazem sentido no vazio – dependem crucialmente *da forma como essas habilidades e competências são medidas*. Qual a complexidade ideal dos textos utilizados para aferir proficiência? Quais as melhores tecnologias para medir essas habilidades? E essas avaliações e tecnologias medem realmente *competência de leitura*?

Este policy review, produzido pelo **Lemann Center** da Stanford Graduate School of Education, discute algumas dessas dimensões, à luz do que fazem 11 outros países e a partir de entrevistas com especialistas nacionais e internacionais.

PRINCIPAIS MENSAGENS:

- As avaliações utilizadas no Brasil para definir parâmetros de proficiência para alfabetização ou fluência leitora nem sempre seguem as melhores práticas mundiais.
- É preciso ir além de habilidades, medindo a capacidade de compreender textos e de fazer conexões entre a leitura e conhecimentos prévios.
- Para tanto, é necessário definir como classificar complexidade textual de forma sistemática, bem como definir a proficiência esperada em diferentes séries para diferentes tipos de texto.
- É possível adaptar avaliações e tecnologias amplamente utilizadas fora do Brasil para avançar nesses desafios.

Como o Brasil mede competência em leitura?

O Saeb é um exame nacional aplicado em diferentes etapas da educação básica (2º, 5º e 9º ano do Ensino Fundamental e 3ª série do Ensino Médio) para medir a proficiência em língua portuguesa e matemática, com frequência a cada dois anos. Os alunos respondem a uma série de perguntas de múltipla escolha e perguntas abertas baseadas em textos fornecidos durante a prova. As provas de Língua Portuguesa do Saeb para o 2º ano são geralmente compostas por tarefas que envolvem ler pequenos textos, identificar letras e palavras, completar frases e associar figuras a palavras. A Avaliação Nacional de Alfabetização (ANA) avaliava os estudantes brasileiros do 3º ano do ensino fundamental (até 2016, tendo abrangência censitária) nas habilidades de leitura, escrita e matemática, com questões de múltipla escolha e questões abertas. Em 2019, a ANA foi descontinuada para dar lugar ao Saeb, quando este passou a avaliar o 2º ano no contexto da BNCC.

Boa parte dos Estados brasileiros possuem sistemas de avaliação de proficiência próprios, tipicamente alinhados com a escala Saeb e aplicados anualmente. Já a mensuração de fluência leitora é mais recente. A Parceria pela Alfabetização em Regime de Colaboração (PARC) passou a medir essa habilidade (no 2º ano do Ensino Fundamental a partir de 2019, em 15 estados brasileiros – Alagoas, Amapá, Espírito Santo, Goiás, Maranhão, Mato Grosso, Mato Grosso do Sul, Pará, Paraíba, Paraná, Pernambuco, Piauí, Rio Grande do Sul, São Paulo e Sergipe –, totalizando 3.271 municípios no país. Nesse contexto, a avaliação de fluência leitora é digital, avaliando precisão, velocidade (incluindo pausas e pontuações) e entoação (ou prosódia). A avaliação é constituída por três partes. Na primeira, os participantes realizam um teste de leitura composto por uma cartela de 60 palavras, dividida em dois blocos. O primeiro bloco contém 20 questões de menor complexidade, enquanto o segundo contém 40 questões com aumento progressivo de complexidade. Na segunda parte, os participantes realizam um teste de leitura de uma cartela composta por 40 *pseudopalavras* – sequências de letras sem significado, com a finalidade de identificar a capacidade de associar sons a sílabas. Na terceira e última parte, os participantes leem um texto narrativo curto (de 110 a 150 palavras) seguido de perguntas interpretativas, com a finalidade de identificar compreensão. São coletados áudios de 60 segundos em cada parte. Ainda que a avaliação seja totalmente digital, apenas a precisão é corrigida automaticamente pelo software. Corretores precisam ouvir os áudios para avaliar ritmo e prosódia, implicando custos financeiros e de tempo expressivos para aplicação dessa maneira de medir fluência leitora, com resultados disponíveis dois meses ou mais depois de sua aplicação. Os modelos de aplicação variam; em alguns Estados, é o próprio professor quem precisa aplicar a prova a cada aluno individualmente de forma sequencial, num processo que consome tempo de instrução e que cria desafios de gestão de sala de aula.

Recentemente, para além dessas avaliações somativas, a PARC passou a realizar também avaliações formativas de fluência leitora em 9 Estados. A Educar para Valer (EpV) realiza essas avaliações formativas de fluência em 48 municípios, seguindo o modelo de Sobral. Essas avaliações são muito mais simples: conduzidas pelo próprio professor como parte regular de suas aulas, sem a ajuda de tecnologias, sem necessidade de correção externa – o próprio professor determina o nível de proficiência a partir de uma rubrica sugerida. A EpV aplica essas avaliações até 6 vezes por ano. Já a PARC, combina avaliações de diferentes naturezas, até 3 vezes por ano: uma diagnóstica no início do letivo, uma formativa no meio, e uma somativa ao final do ano letivo (nessa última, com menos de 15% das matrículas com correção externa; as demais seguem o padrão de correção da formativa).

O que fazem outros países?

A tabela suplementar compila detalhes sobre a mensuração de leitura pelo Brasil e por outros 11 países. Países como Argentina, Alemanha, Paraguai, México, Uruguai, Colômbia e Chile têm focado em avaliações de competência de leitura padronizadas baseadas em questões de múltipla escolha.

Alguns destes, como França, Estados Unidos, Reino Unido, incluem também questões abertas. Na França, o aluno é testado em relação à leitura lendo em voz alta para o professor. Esse formato de avaliação é realizado duas vezes por ano, aplicado pelo próprio professor no contexto da sala de aula, com apoio de rubrica específica – diferente portanto da versão centralizada e digital que o Brasil está tentando implementar. Já nos Estados Unidos, alunos são testados em relação a sua escrita respondendo a questões abertas em formato digital, ou através de tecnologias emergentes para leitura rápida em tela seguida de questões de múltipla escolha – em ambos os casos auto-aplicados pelo aluno e sem necessidade de correção humana subsequente.

Os casos internacionais que mais se aproximam do que vem tentando implementar o Brasil correspondem ao município de Mendoza, na Argentina, e ao Reino Unido até recentemente. No primeiro, o *Censo de Fluidez Lectora* avalia fluência de alunos equivalentes aos 2º a 7º anos do Ensino Fundamental e da 1ª série do Ensino Médio. O método avaliativo se resume à leitura em voz alta em que a professora avalia a performance do aluno ao longo de 60 segundos. O programa atualmente está na sua terceira edição, e objetiva avaliar, anualmente, a precisão, velocidade, e prosódia dos estudantes. Já no Reino Unido, se havia avaliação de fluência nacional com aplicação anual padronizada envolvendo todos os alunos de 5 a 7 anos, esta deixou de ser obrigatória em 2023, na contramão do movimento iniciado recentemente no Brasil.

Vale dizer que todos esses países participam também do *Progress in International Reading Literacy Study* (PIRLS), aplicado em mais de 50 países com o objetivo de avaliar a capacidade de leitura de alunos de 10 anos, passando pela capacidade de recuperar informações, fazer inferências diretas, interpretar e integrar ideias e avaliar conteúdo textual de forma crítica. O Brasil participou do exame pela primeira vez em 2021, tornando possível sua comparação com os demais países participantes.

O que dizem os especialistas?

Medindo alfabetização

No período de 15 a 23 de abril, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) e o MEC realizaram consultas a professoras de todas as regiões do país, a fim de coletar informações sobre o que caracteriza um aluno efetivamente capaz de ler e escrever, de acordo com a BNCC. Em uma segunda fase, em maio, cerca de 20 especialistas em educação discutiram as informações coletadas na etapa anterior e analisaram aspectos técnicos e pedagógicos para, por fim, estabelecerem o critério nacional de alfabetização. Esse é um passo de extrema importância e um avanço significativo. Primeiro, porque torna transparente o que se espera de uma criança ao final do 2º ano no que diz respeito à alfabetização. Segundo, porque finalmente temos um critério oficial que deve servir como base para que os diferentes Estados e municípios possam desenvolver suas políticas de alfabetização com diagnósticos e metas mais precisos.

Mais importante que a definição do parâmetro de 743 pontos no Saeb é a definição do que significa, de fato, uma criança estar alfabetizada no país. O avanço está em estabelecer com clareza o conceito de criança alfabetizada no contexto nacional, levando em consideração a contribuição de professores alfabetizadoras e especialistas na área. Esse é um passo crucial.

Os desafios residem na aplicação prática dessa visão: que tipo de texto e de complexidade devem ser lidos pelas crianças do 2º ano? O que caracteriza uma criança que possui habilidades de leitura e escrita? Quais erros ortográficos são considerados esperados nessa etapa? O que significa ser um leitor iniciante? Esses elementos são essenciais para auxiliar os professores, gestores escolares e gestores públicos a terem uma compreensão concreta que possa apoiar os estudantes em seu processo de aprendizagem.

Além de enfrentar esses desafios, é igualmente importante, agora que há uma visão consolidada no país, comparar essa visão com padrões internacionais, a fim de garantir altas expectativas para nossas crianças. É fundamental também articular o Saeb com outros modelos de avaliação, especialmente aqueles relacionados à leitura, e integrar esses instrumentos à visão estabelecida, uma vez que o Saeb não mensura de forma abrangente algumas das competências mencionadas nessa perspectiva. Por fim, é necessário tornar transparente a progressão esperada nos anos iniciais do Ensino Fundamental, para que as crianças possam avançar além do nível básico.

O problema competência de leitura

Francisco (Chico) Soares, professor de Educação na UFMG, se volta para a discussão de mensuração de competência de leitura para além de habilidades específicas. Escreve: “Recentemente foram divulgados os resultados do Estudo Internacional de Progresso em Leitura (PIRLS), uma avaliação internacional que mede o nível da compreensão de textos dos estudantes de 10 anos em muitos países do mundo. As proficiências dos estudantes brasileiros foram especialmente baixas.”



Chico Soares
Professor de Educação
na UFMG

“As explicações certamente mostrarão o impacto das condições sociais dos estudantes, das condições de funcionamento das escolas e da formação dos docentes, do envolvimento dos estudantes com o teste. Sem minimizar a importância destas explicações, busco contribuir para o debate salientando que o resultado é também consequência das escolhas pedagógicas feitas pelos sistemas de ensino no Brasil, diferentes daquelas usadas pelo PIRLS.”

Chico continua: “O instrumento usado pelo PIRLS para medir a proficiência dos estudantes na compreensão de textos é composto de apenas dois textos, um literário e outro informativo, cada um com aproximadamente 15 perguntas. Estas perguntas, apresentadas tanto como questões de múltipla escolha como questões de resposta construída, no seu conjunto, geram a informação considerada *suficiente* para a estimação da proficiência em compreensão de textos de cada estudante. O fato de o instrumento conter muitas questões sobre um mesmo texto é crucial para que se possa verificar o domínio do estudante na competência leitora, capacidade que exige a mobilização concomitante de várias habilidades de leitura e conhecimentos linguísticos. Importante ressaltar que os textos usados no PIRLS são textos autênticos, com complexidade adequada para a etapa de vida dos estudantes. O conceito de complexidade inclui, mas não se restringe, ao tamanho do texto.

O modelo conceitual do PIRLS sugere que o projeto pedagógico de uma rede de ensino deveria mostrar quais textos os estudantes devem ser capazes de compreender ao fim de cada etapa de ensino. A verificação do desenvolvimento da capacidade de compreensão de textos exige o uso de um conjunto de questões que, no seu conjunto, geram evidência adequada dessa competência. Estas questões precisam contemplar as diferentes complexidades cognitivas exigidas. O PIRLS trabalha com quatro categorias: (i) localizar e recuperar informações explícitas; (ii) fazer inferências diretas;

(iii) interpretar e articular ideias e informações; e (iv) avaliar e analisar criticamente o conteúdo e os elementos textuais. Questões sobre as duas últimas categorias são especialmente necessárias para a verificação da compreensão do texto como um todo já que sua resposta adequada exigem a compreensão global do texto, não apenas o uso de informação localizada em partes específicas. Estas questões são mais difíceis de serem formuladas como múltipla escolha. Isso explica o uso de respostas construídas pelo PIRLS.

Esta estrutura não é a utilizada no SAEB, avaliação cujo instrumento é organizado de forma que cada questão visa verificar se o estudante desenvolveu uma habilidade leitora específica. Por isso, o suporte textual destas questões são recortes de textos que contêm apenas a informação necessária para a resposta da questão específica. Assim sendo, o estudante brasileiro está acostumado a ler textos de complexidade muito mais baixa do que seria o adequado para o seu ano escolar e, portanto não desenvolve a capacidade de mobilizar as várias habilidades necessárias para compreensão global do texto.

Esta diferença de modelos conceituais leva a testes diferentes e, naturalmente, a diagnósticos diferentes. Por exemplo, os últimos resultados do IDEB para o 5º ano, sinalizaram sucesso nesta etapa do ensino. Isso tem apoiado o diagnóstico de que o problema do ensino brasileiro se concentra no Fundamental II. Os resultados do PIRLS sugerem, no entanto, que os estudantes que chegam ao Fundamental II não são capazes de ler o que precisam para desenvolver os conhecimentos esperados para esta etapa.”

A questão da complexidade textual

Continua Chico: “O programa pedagógico de ensinar leitura através de textos já circula no Brasil há muitos anos. A saudosa Magda Soares deu o nome de ‘Português através de textos’ à sua primeira coleção de livros didáticos para o Ensino Fundamental. A consulta a estes livros mostra como o programa do PIRLS já estava completamente presente no debate pedagógico brasileiro. O cuidado na seleção dos textos e abrangência da complexidade cognitiva das questões são marcas deste material. Também a BNCC faz muitas menções a escolha de textos adequados para a faixa etária (embora não defina o que isso signifique na prática). Por exemplo a resolução do CNE que instituiu da BNCC diz que a alfabetização deve garantir [...] aos estudantes a apropriação do sistema de escrita alfabética, a compreensão leitora e a escrita de textos com complexidade adequada à faixa etária dos estudantes.’

Infelizmente a centralidade do texto para o ensino e avaliação se perdeu nos últimos anos. Recentemente, os Estados e municípios construíram seus currículos a partir da BNCC, mas optaram por fazer mera repetições das habilidades de leitura, não enfatizando a escolha de textos em que estas habilidades seriam verificadas e desenvolvidas. Seria de se esperar que o processo de construção curricular identificasse textos culturalmente relevantes para os estudantes das diferentes regiões do Brasil e com a complexidade adequada a cada ano escolar.

O descompasso entre o modelo conceitual usado no Brasil e nas referências internacionais que se observa em leitura ocorre também em matemática. Os testes brasileiros são constituídos de questões de baixa complexidade cognitiva e não contemplam a verificação - essencial - da capacidade de o estudante mobilizar diferentes habilidades para a solução de problemas, exatamente o que a vida lhes exige.”

Medindo fluência leitora

No que diz respeito à maneira como o Brasil mede fluência leitora, Guilherme Lichand, professor de Educação na Stanford GSE, questiona: “A capacidade de vocalizar um texto com precisão, ritmo e prosódia é, de fato, saber ler? Mesmo que correlacionada, a verdade é que fluência leitora não é condição necessária nem suficiente para competência de leitura. Não é necessária porque é possível saber ler com competência sem conseguir vocalizar – basta

pensar, por exemplo, num aluno mudo. Tampouco é suficiente porque é possível vocalizar com perfeição sem entender o significado de uma única frase que foi lida.”

O laboratório de leitura e transtornos de aprendizagem da Universidade de Stanford tem uma abordagem bem diferente para medir essa competência. O *Rapid Online Assessment of Reading* (ROAR) tem diferentes versões. Na mais simples, sequências de palavras piscam na tela por poucos segundos. Se o aluno lê com fluidez, é capaz de identificar corretamente na pergunta subsequente a alternativa com a palavra que piscou. Em outras versões do teste, o aluno escuta palavras ou frases, e tem que identificar na tela do computador, tablet ou celular o texto correspondente. As alternativas apresentadas nas questões de múltipla escolha podem ainda envolver sinônimos e até exigir que o aluno faça inferência a partir do texto que teve que ler no tempo disponível.

Saber identificar as palavras ou frases que escutou no tempo designado requer competência leitora. É, portanto, um teste de competência, para além de habilidades como prosódia ou pontuação. Como benefícios conjugados, o aluno se auto aplica o teste – sem necessidade de mediação do professor –, o software corrige tudo automaticamente e o resultado fica disponível na hora. Jason Yeatman, professor de Educação da Stanford GSE, aponta que o ROAR avalia distritos escolares inteiros em 20 minutos e sem custos.

Mais do que isso, escolhendo opções de resposta estrategicamente, é possível usar o sistema para obter sinais de transtornos de aprendizagem como dislexia, o que não é possível fazer na metodologia que começou a ser escalada nos Estados brasileiros.



Jason Yeatman
Professor de Educação
na Stanford GSE

“As evidências científicas apontam que a metodologia do ROAR não só é capaz de medir as mesmas habilidades que as aplicações que envolvem vocalização de textos procuram capturar, mas ainda permite ir além – medindo outras habilidades, competências e transtornos de aprendizagem, além de ser muito mais simples de aplicar, rápida, gratuita e de código aberto.”

Avaliações somativas baseadas em metodologia como a do ROAR podem ser aplicadas múltiplas vezes por ano, em função de sua simplicidade, rapidez e custo (baixo ou zero).

Essas avaliações podem ser complementadas por avaliações formativas, realizadas pelo próprio professor com apoio de rubrica específica, tão frequentemente quanto uma vez por mês. Luciana Alves, professora de Fonoaudiologia da UFMG, destaca experiências bem-sucedidas de seu grupo no Brasil com rubricas simplificadas – de modo a garantir que o professor saiba como classificar os alunos com base na proficiência esperada para cada conjunto de habilidades que compõem a fluência leitora, e que essa classificação seja feita de forma homogênea por diferentes professores.



Luciana Mendonça Alves
Professora de Fonoaudiologia
na UFMG

“O importante é que próprio professor possa aplicar essa avaliação por diversas vezes ao longo do ano escolar. O filme é mais revelador do que a fotografia. Os alunos com dificuldades não são necessariamente aqueles que começam sub-proficientes, mas os que não evoluem em relação a si mesmos ao longo do ano escolar.”

A Escala de Percepção e Análise da Fluência Leitora (SOLAR), rubrica para classificação de fluência leitora pelo próprio professor a partir de excertos associados a diferentes graus de complexidade textual, desenvolvido pelo grupo de Luciana em parceria com a Professora da UnB Letícia Celeste, é uma referência promissora que poderia ser escalada para todo o país.

Chico Soares reforça que “a fluência é um resultado intermediário muito importante; já que não há compreensão sem fluência adequada. Assim medir este resultado durante o processo de alfabetização, por exemplo na educação infantil e no primeiro ano do Fundamental, pode ajudar. A fluência, entretanto, *não é o resultado final* – que deveria ser a capacidade de compreender textos com complexidade adequada para a idade dos estudantes.”

Sobre o formato dos testes, o professor ainda destaca que “Os testes internacionais como PIRLS e PISA usam tanto itens de múltipla escolha como de resposta construída. Verificar o desenvolvimento de habilidades de ordem cognitiva mais alta é muito mais difícil com testes de múltipla escolha. O problema não é, intrinsecamente, com a forma, mas as consequências de uso exclusivo de uma única forma. A experiência internacional que usa questões dos dois formatos, deveria nos ensinar algo. No fundo esta discussão é apenas parte da questão maior dos objetivos da avaliação. Se for apenas para o monitoramento o tipo de questão importa menos. Se for para subsidiar ações pedagógicas, as questões de múltipla escolha não conseguem produzir indicações mais sólidas.”

Se pode ser mais caro coletar e corrigir respostas abertas, o professor aponta que não é preciso aplicar essas questões ao universo dos alunos da rede para uma avaliação somativa capaz de informar políticas da rede. “A resposta para isso chama-se amostragem. Dito isso, é preciso gastar um pouco dos substanciais recursos atualmente alocados à avaliação para avanços tecnológicos que permitam o uso mais fácil de questões abertas. Os recentes avanços do processamento de linguagem natural (*Natural Language Processing*, NLP) indicam que isso é possível.”

Chico reforça a relevância complementar de avaliações formativas, realizados pelos próprios professores na sala de aula, de forma descentralizada, mas alerta sobre a necessidade de diretrizes orientadoras. “No caso de fluência, a escolha das palavras é fundamental. A função do Estado seria providenciar conjuntos de palavras que pudessem ser usadas nas diferentes etapas do aprendizado. Em português a estrutura silábica padrão é CV – consoante vocal. Mas há outras estruturas. Há palavras mais raras, menos concretas. Ou seja, um conhecimento já sistematizado que deve ser colocado à disposição das professoras. Eu penso que isso deve ser feito no âmbito da escola e, depois, da rede.”

O que diz a Stanford LC Expert Network



Filomena Siqueira
Gerente Pedagógica
no Instituto Reúna

“Em termos de abordagem, precisamos superar as disputas entre métodos, que só cria manchetes e alimenta posições políticas, mas em nada contribui para o trabalho do docente e aprendizagem dos estudantes. Sabemos que, sem a capacidade de decodificar, não será possível para os estudantes adquirir fluência leitora e, assim, conseguir compreender e interpretar textos. Superada essa dicotomia improdutiva, práticas em sala de aula como leitura oral e redação são preditores importantes de alfabetização e devem ser realizados frequentemente durante as aulas.”



Fernando Tavares Jr
Pesquisador na
UFJF - CAEd

“Ainda são mobilizadas poucas estratégias que se fazem mais uso das tecnologias disponíveis, como apps disponíveis para uso em vários dispositivos (i.e., PlayStore), AI aplicada à leitura, avaliações adaptativas (i.e., *gaming*) e outras similares. A avaliação da leitura ainda é muito centrada na tradição das letras, deixando de lado outros símbolos e signos que se mostram relevantes. A relação entre uma imagem e a identificação da informação expressa nela é fundamental para se compreender a multidimensionalidade da competência leitora.”



Fernanda A. Yamamoto
Coordenadora de Projetos
Educacionais no SENAC

“Elaboração de indicadores de desempenho específicos e que sejam trabalhados em cursos de formação continuada de professores, cujo desenvolvimento e capacidade de avaliação neste contexto são primordiais. Dessa forma, estaríamos tornando possível a revisão de estratégias e metodologias utilizadas em sala de aula.”



Gabriela Fiore Bonicio
Assessment Coordinator
na Camino School

“Precisamos rever nossas avaliações de larga escala. O formato de múltipla escolha captura habilidades rasas de leitura e é insuficiente para informar escolas e professores sobre próximos passos para impulsionar a aprendizagem de forma equitativa. A Austrália, por exemplo, já usa instrumento diversos como respostas abertas, redação e compreensão oral. Além disso, usa tecnologia para adaptar a dificuldade de suas questões e fornecer informações detalhadas ao nível societário, escolar e individual sobre o desempenho de seus estudantes.”