

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

Análise da Dificuldade de Palavras em Avaliações de Fluência Leitora Utilizando a Teoria de Resposta ao Item

Ary de Paula Canuto Neto

JUIZ DE FORA
JULHO, 2026

Análise da Dificuldade de Palavras em Avaliações de Fluência Leitora Utilizando a Teoria de Resposta ao Item

ARY DE PAULA CANUTO NETO

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Bacharelado em Ciência da Computação

Orientador: Jairo Francisco de Souza

JUIZ DE FORA

JULHO, 2026

ANÁLISE DA DIFICULDADE DE PALAVRAS EM AVALIAÇÕES DE FLUÊNCIA LEITORA UTILIZANDO A TEORIA DE RESPOSTA AO ITEM

Ary de Paula Canuto Neto

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS
EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTE-
GRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE
BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

Jairo Francisco de Souza
Doutor em Informática - PUC-Rio

Heder Soares Bernardino
Doutor em Modelagem Computacional - LNCC

Wellington Silva
Doutor em Educação - PUC-Rio

JUIZ DE FORA
13 DE JULHO, 2026

Resumo

Este trabalho investiga a dificuldade de leitura de palavras isoladas em língua portuguesa por meio da Teoria de Resposta ao Item (TRI), utilizando dados das avaliações PARC Entrada 2024 e PARC Saída 2024. Após o tratamento e a codificação das respostas dos estudantes, foi aplicado o Modelo Logístico Unidimensional de 1 Parâmetro (Modelo de Rasch) para estimar a proficiência dos participantes e a dificuldade dos itens. Inicialmente, foram verificadas as premissas do modelo por meio da análise de unidimensionalidade. Em seguida, analisou-se a distribuição da dificuldade das palavras nos diferentes cadernos de prova, demonstrando equilíbrio entre eles apesar da variação na ordem de apresentação dos itens. Também foi comparada a proficiência estimada pela TRI com a quantidade de palavras corretas, evidenciando forte associação entre ambas as medidas. Adicionalmente, investigou-se a influência de características linguísticas e de frequência de uso das palavras sobre sua dificuldade por meio de modelos de regressão linear múltipla, identificando fatores relevantes para sua predição. Por fim, verificou-se que as dificuldades estimadas pela regressão preservam, em grande medida, as estimativas de proficiência obtidas pela TRI, indicando que características linguísticas podem contribuir para a construção e o aprimoramento de instrumentos de avaliação da fluência leitora.

Palavras-chave: Teoria de Resposta ao Item, Modelo de Rasch, Fluência Leitora, Leitura de Palavras, Avaliação Educacional.

Abstract

This study investigates the reading difficulty of isolated words in Brazilian Portuguese using Item Response Theory (IRT), based on data from the PARC Entry 2024 and PARC Exit 2024 assessments. After preprocessing and coding the students' responses, the One-Parameter Logistic Model (Rasch Model) was applied to estimate students' proficiency and item difficulty. First, the assumptions of the model were verified through a unidimensionality analysis. Subsequently, the distribution of word difficulty across different test booklets was examined, demonstrating a balanced distribution despite variations in the order of item presentation. Students' proficiency estimates obtained through IRT were also compared with the traditional measure based on the number of correctly read words, revealing a strong association between the two metrics. Additionally, the influence of linguistic characteristics and word frequency on item difficulty was investigated using multiple linear regression models, identifying relevant factors for predicting word difficulty. Finally, the results showed that the difficulties predicted by the regression model largely preserve the proficiency estimates obtained through IRT, indicating that linguistic features can contribute to the development and improvement of reading fluency assessment instruments.

Keywords: Item Response Theory, Rasch Model, Reading Fluency, Word Reading, Educational Assessment.

Agradecimentos

Aos meus pais Alexandre e Rosaria e à minha irmã Carol, por todo o apoio incondicional durante a graduação.

Ao meu amigo Thiago, que nunca deixou de estar presente e me apoiar em todos os momentos.

À minha amiga Jhulia, por sempre ouvir meus devaneios e não desistir de mim.

À minha amiga Mariana, por sua presença que, em muitos momentos, me trouxe tranquilidade.

Ao professor Jairo, por me acolher como seu orientando e por despertar em mim o interesse por uma área pela qual tive grande satisfação em aprender.

E, por fim, a todos os meus amigos que fizeram parte da minha jornada na graduação.

Conteúdo

Lista de Figuras	5
Lista de Tabelas	6
Lista de Abreviações	7
1 Introdução	8
1.1 Descrição do Problema	9
1.2 Justificativa	10
1.3 Objetivos	11
2 Fundamentação Teórica	13
2.1 Leitura e Reconhecimento de Palavras	13
2.2 Psicometria	15
2.2.1 Modelo Logístico Unidimensional de 1 Parâmetro	17
2.2.2 Representações Gráficas para Análise Estatística	18
3 Trabalhos Relacionados	20
3.1 Avaliação da Leitura de Palavras e Desenvolvimento de Instrumentos	20
3.2 Aplicação da Teoria de Resposta ao Item em Avaliações de Leitura	21
3.3 Características Linguísticas Associadas à Dificuldade de Leitura	22
3.4 Análise Comparativa	23
3.5 Considerações Finais	24
4 Metodologia	26
4.1 Caracterização da pesquisa	26
4.2 Caracterização da avaliação de leitura de palavras	26
4.3 Descrição das bases de dados	27
4.4 Codificação das respostas	29
4.5 Aplicação da Teoria de Resposta ao Item	29
5 Análise dos resultados	32
5.1 Verificação da Unidimensionalidade	32
5.2 Distribuição da Dificuldade dos Cadernos	35
5.2.1 Distribuição Acumulada da Dificuldade dos Cadernos	36
5.2.2 Dificuldade dos Itens por Posição	38
5.3 Distribuição da Proficiência dos Estudantes	39
5.3.1 Comparação entre PARC Entrada 2024 e PARC Saída 2024	40
5.4 Análise da Dificuldade das Palavras	45
5.4.1 Estimativa da Proficiência dos Estudantes a partir das Dificuldades Preditas	49
5.5 Discussão	52
6 Conclusões	55
Bibliografia	57

Lista de Figuras

5.1	Gráficos de sedimentação.	34
5.2	Distribuição acumulada da dificuldade dos cadernos da avaliação PARC Entrada 2024	37
5.3	Distribuição acumulada da dificuldade dos cadernos da avaliação PARC Saída 2024	38
5.4	Distribuição da dificuldade dos itens por posição nos cadernos da avaliação PARC Entrada 2024	39
5.5	Distribuição da dificuldade dos itens por posição nos cadernos da avaliação PARC Saída 2024	39
5.6	Distribuição da proficiência estimada pela TRI. Os gráficos utilizam escalas independentes nos eixos horizontal e vertical para facilitar a visualização das distribuições.	40
5.7	Distribuição de quantidade de palavras corretas. Os gráficos utilizam es- calas independentes no eixo vertical para facilitar a visualização das distri- buições.	41
5.8	Distribuição percentual dos estudantes e dos itens por faixas da escala da TRI nas avaliações PARC Entrada 2024 e PARC Saída 2024. As faixas foram definidas a partir da média e do desvio-padrão das proficiências es- timadas em cada aplicação.	42
5.9	Relação entre QPC e proficiência TRI. Os gráficos utilizam escalas inde- pendentes no eixo vertical para facilitar a visualização das distribuições. . .	43
5.10	Proficiência média estimada pela TRI em função da quantidade de palavras corretas (QPC). Os gráficos utilizam escalas independentes no eixo vertical para facilitar a visualização das distribuições.	44
5.11	Dificuldade prevista X calculada pela TRI.	46
5.12	Comparação entre a proficiência estimada pela TRI e a proficiência esti- mada utilizando as dificuldades previstas pelo modelo de regressão linear múltipla.	50

Lista de Tabelas

3.1	Tabela comparativa entre trabalhos relacionados	24
4.1	Caracterização das avaliações analisadas	28
4.2	Codificação preliminar das respostas	29
5.1	Resultados da análise de dimensionalidade – PARC Entrada 2024	33
5.2	Resultados da análise de dimensionalidade – PARC Saída 2024	34
5.3	Resultados da correlação e da regressão linear simples entre QPC e proficiência estimada	44
5.4	Variáveis independentes utilizadas no modelo de regressão linear múltipla.	46
5.5	Comparação entre os modelos de regressão linear múltipla para a PARC Entrada e PARC Saída	47
5.6	Comparação dos coeficientes estimados para os modelos de regressão linear múltipla	48
5.7	Comparação entre as proficiências estimadas pela TRI e pela regressão linear múltipla	51

Lista de Abreviações

QPC	Quantidade de Palavras Corretas
TRI	Teoria de Resposta ao Item
BNCC	Base Nacional Comum Curricular
ENEM	Exame Nacional do Ensino Médio
TDE	Teste de Desempenho Escolar
PLEP	Prova de Leitura e Escrita de Palavras
d-ELP	Developmental English Lexicon Project
PARC	Parceria pela Alfabetização em Regime de Colaboração

1 Introdução

A avaliação educacional constitui um dos principais instrumentos utilizados para acompanhar, compreender e mensurar os processos de ensino e de aprendizagem. Sua função ultrapassa a simples atribuição de notas ou a classificação de estudantes, assumindo um papel relevante na identificação de dificuldades de aprendizagem, no monitoramento do desempenho escolar e no suporte à tomada de decisões pedagógicas e institucionais (BONAMINO; SOUSA, 2012).

Historicamente, os processos avaliativos estiveram associados à verificação quantitativa do rendimento escolar, sendo frequentemente utilizados como mecanismos de aprovação, reprovação e seleção acadêmica (DANTAS, 2023). Entretanto, o desenvolvimento das pesquisas em educação contribuiu para ampliar a compreensão da avaliação, que passou a ser entendida como componente integrante do processo educativo (BLOOM; HASTINGS; MADAUS, 1971).

Além de seu papel pedagógico, a avaliação educacional passou a desempenhar uma função estratégica na formulação e no monitoramento de políticas públicas educacionais. A partir da década de 1990, diversos países latino-americanos ampliaram o uso de avaliações em larga escala como forma de produzir indicadores da qualidade da educação e monitorar o desempenho dos sistemas educacionais (TEDESCO, 2003).

No Brasil, a primeira experiência foi com o Sistema de Avaliação da Educação Básica (SAEB), que passou a desempenhar um papel central na produção de indicadores educacionais. Esse sistema possibilita acompanhar o desempenho dos estudantes em diferentes etapas da escolarização, além de fornecer subsídios para a elaboração de políticas públicas voltadas à melhoria da qualidade da educação (BECKER, 2010).

No contexto brasileiro, iniciativas voltadas à avaliação da fluência leitora em larga escala ainda são relativamente recentes. Um exemplo é o estudo desenvolvido por Silva, Silva e Megherbi (2023), que descreve a construção e a validação de um instrumento de avaliação da fluência em leitura aplicado a estudantes do Ensino Fundamental em diferentes regiões do país. Os autores propõem um modelo de avaliação baseado na leitura

de palavras, pseudopalavras e textos, fundamentado em evidências de validade e fidedignidade, demonstrando o potencial desse tipo de instrumento para subsidiar o diagnóstico da alfabetização e orientar intervenções pedagógicas. Essa experiência evidencia o crescente interesse pelo desenvolvimento de avaliações capazes de mensurar, de forma objetiva e confiável, habilidades relacionadas à leitura, reforçando a relevância de pesquisas que busquem compreender o comportamento psicométrico dos itens que compõem esses instrumentos.

A expansão das avaliações educacionais em larga escala intensificou a necessidade de desenvolver instrumentos capazes de mensurar com maior precisão a aprendizagem dos estudantes. Durante muitos anos, os sistemas educacionais concentraram-se principalmente na produção de estatísticas sobre o funcionamento da educação, sem avaliar efetivamente o que os alunos aprendiam. Entretanto, a percepção de baixos níveis de aprendizagem e o desafio de expandir simultaneamente o acesso e a qualidade da educação impulsionaram mudanças no enfoque das avaliações educacionais, favorecendo a adoção de metodologias estatísticas e psicométricas mais sofisticadas (BECKER, 2010)

1.1 Descrição do Problema

A avaliação da leitura de palavras isoladas é amplamente utilizada em contextos educacionais para mensurar habilidades relacionadas à alfabetização e ao desenvolvimento da fluência leitora. Para que esses instrumentos sejam capazes de diferenciar estudantes com distintos níveis de proficiência, é necessário que as palavras utilizadas apresentem diferentes níveis de dificuldade.

A determinação da dificuldade de palavras é tradicionalmente realizada com base em critérios linguísticos e psicolinguísticos, como número de sílabas, frequência de ocorrência, regularidade ortográfica e complexidade fonológica. Entretanto, esses critérios nem sempre refletem com precisão a dificuldade observada na prática, uma vez que o desempenho dos leitores pode ser influenciado por diversos fatores cognitivos e linguísticos. Dessa forma, a estimativa da dificuldade de palavras representa um desafio para a construção de instrumentos de avaliação da leitura.

Com a crescente disponibilidade de bases de dados contendo respostas de es-

tudantes em avaliações educacionais, torna-se possível utilizar métodos estatísticos para estimar empiricamente a dificuldade dos itens a partir do desempenho efetivamente observado. Entre esses métodos, a Teoria de Resposta ao Item (TRI) destaca-se por permitir a modelagem da relação entre a habilidade dos participantes e a probabilidade de acerto dos itens, produzindo estimativas mais objetivas de dificuldade.

Este trabalho propõe a utilização da Teoria de Resposta ao Item (TRI), por meio do Modelo de Rasch, para estimar a dificuldade de leitura de palavras isoladas em língua portuguesa a partir de uma base de respostas de estudantes. Além da estimação dos parâmetros de dificuldade dos itens e da proficiência dos participantes, são realizadas análises para verificar as premissas do modelo, avaliar a distribuição da dificuldade entre diferentes cadernos de prova e investigar a relação entre características linguísticas das palavras, sua frequência de uso e as dificuldades estimadas.

A partir dessas análises, busca-se produzir evidências empíricas sobre os fatores associados à dificuldade de leitura de palavras isoladas em língua portuguesa. Espera-se que os resultados contribuam para a construção e o aprimoramento de instrumentos de avaliação da fluência leitora, além de fornecer subsídios para pesquisas futuras sobre avaliação educacional e reconhecimento de palavras.

1.2 Justificativa

A estimação da dificuldade de palavras para avaliações de leitura apresenta diversas justificativas relevantes. Uma das principais é a possibilidade de produzir estimativas empíricas da dificuldade dos itens com base no desempenho observado dos estudantes, complementando os critérios linguísticos tradicionalmente utilizados na construção de instrumentos de avaliação.

Ao estimar a dificuldade das palavras por meio da Teoria de Resposta ao Item (TRI), torna-se possível compreender com maior precisão como cada item se comporta diante de diferentes níveis de proficiência leitora. Essa abordagem permite identificar palavras mais fáceis e mais difíceis a partir de evidências empíricas, contribuindo para a construção de testes mais equilibrados e capazes de discriminar adequadamente estudantes com diferentes habilidades de leitura.

Além disso, a utilização de uma ampla base de respostas possibilita obter estimativas fundamentadas em dados observados, aumentando a confiabilidade das análises. Essas estimativas podem auxiliar pesquisadores e profissionais da educação na elaboração de novos instrumentos, na seleção de itens e na calibração de avaliações de leitura.

Outro aspecto relevante consiste na investigação da relação entre características linguísticas, frequência de uso das palavras e os níveis de dificuldade estimados. Essa análise permite identificar quais fatores exercem maior influência sobre a dificuldade de leitura e avaliar em que medida essas características podem ser utilizadas para prever a dificuldade de novos itens.

De maneira geral, a utilização da TRI, aliada à análise de características linguísticas das palavras, contribui tanto para o avanço das pesquisas em avaliação educacional quanto para o desenvolvimento de instrumentos de avaliação da fluência leitora mais fundamentados em evidências empíricas.

1.3 Objetivos

O objetivo deste trabalho é utilizar a Teoria de Resposta ao Item (TRI) para estimar a dificuldade de leitura de palavras isoladas em língua portuguesa a partir de uma base de respostas de estudantes. Busca-se produzir uma escala empírica de dificuldade dos itens, fundamentada no desempenho observado dos participantes, que possa servir como referência para a construção e análise de instrumentos de avaliação da leitura.

Para atingir esse objetivo, são definidos os seguintes passos:

- Aplicar um modelo da Teoria de Resposta ao Item para estimar os parâmetros de dificuldade das palavras presentes na avaliação.
- Analisar a distribuição das dificuldades estimadas, identificando palavras com diferentes níveis de complexidade para os leitores.
- Investigar a relação entre características linguísticas das palavras e os níveis de dificuldade obtidos por meio da TRI.
- Analisar a distribuição de proficiência dos estudantes estimada pelo modelo e com-

parar esses resultados com indicadores tradicionais de desempenho baseados na quantidade de palavras lidas corretamente.

- Disponibilizar evidências empíricas sobre a dificuldade de leitura de palavras isoladas em língua portuguesa, contribuindo para pesquisas futuras na área de avaliação educacional e fluência leitora.

Este trabalho contribui para a literatura ao produzir estimativas empíricas de dificuldade de palavras em língua portuguesa com base em uma ampla amostra de leitores e utilizando a Teoria de Resposta ao Item. Além disso, os resultados geram uma referência que pode auxiliar pesquisadores e profissionais envolvidos na construção de instrumentos de avaliação da leitura.

2 Fundamentação Teórica

2.1 Leitura e Reconhecimento de Palavras

A aprendizagem da leitura constitui um processo complexo que envolve o desenvolvimento gradual da capacidade de relacionar os símbolos gráficos da escrita aos sons da fala. Entre os modelos teóricos que buscam explicar esse processo, destaca-se a teoria proposta por Ehri (2005), segundo a qual a leitura eficiente depende da formação de conexões entre grafemas e fonemas, permitindo que a representação ortográfica de uma palavra seja associada à sua pronúncia e ao seu significado na memória.

De acordo com essa perspectiva, a aprendizagem da leitura não ocorre de forma instantânea, mas por meio de um processo gradual de desenvolvimento. Inicialmente, as crianças encontram-se na fase pré-alfabética, na qual reconhecem palavras utilizando pistas visuais gerais, como formato, tamanho ou elementos gráficos salientes, sem compreender efetivamente as relações entre letras e sons. Em seguida, ocorre a fase alfabética parcial, caracterizada pela utilização de algumas pistas fonéticas, geralmente relacionadas às letras iniciais ou finais das palavras. Posteriormente, os estudantes atingem a fase alfabética completa, momento em que passam a utilizar sistematicamente as correspondências entre grafemas e fonemas para realizar a decodificação das palavras. Finalmente, desenvolve-se a fase alfabética consolidada, na qual o leitor passa a reconhecer padrões ortográficos maiores, como sílabas, morfemas e sequências recorrentes de letras, tornando o reconhecimento das palavras mais rápido e eficiente.

O resultado desse processo é o desenvolvimento da chamada leitura por reconhecimento imediato (*sight word reading*). Nessa condição, a identificação das palavras ocorre de forma praticamente automática, sem a necessidade de decodificação consciente de cada letra ou sílaba. Segundo Ehri (2005), quanto maior o número de palavras armazenadas na memória ortográfica do leitor, menor o esforço cognitivo necessário para seu reconhecimento. Consequentemente, mais recursos atencionais podem ser direcionados para atividades de nível superior, como a compreensão e a interpretação do texto. Essa

capacidade de reconhecimento automático é considerada um dos principais componentes da fluência em leitura e um importante indicador do desenvolvimento da alfabetização.

Embora os princípios gerais do desenvolvimento da leitura sejam compartilhados entre diferentes idiomas, as características de cada sistema de escrita influenciam diretamente esse processo. Conforme observado por Seymour, Aro e Erskine (2003), o português brasileiro apresenta menor complexidade ortográfica e predominância de sílabas do tipo Consoante-Vogal (CV) quando comparado ao inglês, favorecendo a aquisição inicial da decodificação. Entretanto, o desenvolvimento da leitura em português também exige a consolidação do reconhecimento ortográfico, permitindo que as palavras sejam identificadas de forma rápida e automática.

Nesse contexto, a sílaba assume papel central no processo de alfabetização em língua portuguesa. Segundo Hachén (2002), a percepção e o reconhecimento de unidades silábicas precedem a análise fonêmica, pois as sílabas constituem unidades linguísticas mais acessíveis às crianças do que os fonemas isolados. Assim, a estrutura Consoante-Vogal (CV), predominante na língua portuguesa, funciona como referência para o desenvolvimento inicial das habilidades de leitura.

À medida que o processo de aprendizagem avança, estruturas silábicas mais complexas, como V, CVC, CCV e CVV, passam a ser introduzidas (Brasil. Ministério da Educação, 2018). A transição para esses padrões exige que o estudante desenvolva maior capacidade de análise fonológica e ortográfica, identificando segmentos em posições de coda ou ataque complexo. Conforme demonstrado por Seymour, Aro e Erskine (2003), o aumento da complexidade silábica demanda maior processamento cognitivo durante a decodificação, fazendo com que palavras contendo encontros consonantais, dígrafos ou marcas de nasalidade tendam a apresentar maior dificuldade de leitura.

A compreensão desses processos cognitivos e linguísticos fundamenta diretamente a construção de instrumentos destinados à avaliação da leitura. Em avaliações de alfabetização, a unidade básica de mensuração é o item, geralmente composto por uma instrução e um estímulo, que pode ser representado por letras, sílabas, palavras ou pseudopalavras. Segundo Dubeck e Gove (2015), esses itens devem ser elaborados de forma a medir habilidades específicas relacionadas ao desenvolvimento da leitura, permitindo identificar o

nível de domínio do estudante sobre os processos de decodificação e reconhecimento de palavras.

Além dos aspectos cognitivos envolvidos na aprendizagem, a elaboração dos itens deve considerar características linguísticas capazes de influenciar sua dificuldade. Conforme discutido por Ehri (2005), leitores em fases iniciais apoiam-se predominantemente em pistas visuais e fonológicas parciais, tornando importante controlar a similaridade entre os estímulos utilizados nas avaliações. Da mesma forma, Dubeck e Gove (2015) destacam que a utilização de palavras familiares ao vocabulário infantil reduz a influência do conhecimento lexical sobre o desempenho, permitindo que o instrumento avalie prioritariamente a habilidade de leitura.

No contexto da língua portuguesa, a Base Nacional Comum Curricular estabelece que o ensino e a avaliação da leitura devem partir de estruturas silábicas mais simples e regulares, introduzindo progressivamente padrões de maior complexidade, como sílabas CVC e CCV, encontros consonantais, dígrafos e marcas de nasalidade (Brasil. Ministério da Educação, 2018). Dessa forma, características como quantidade de sílabas, padrão silábico, canonicidade, frequência de ocorrência e familiaridade lexical constituem variáveis relevantes para a seleção de palavras e para a determinação da dificuldade dos itens em avaliações de leitura.

Em síntese, a construção de instrumentos de avaliação da leitura fundamenta-se na integração de conhecimentos oriundos da psicologia cognitiva, da linguística e da psicometria. Enquanto os modelos cognitivos descrevem como ocorre o desenvolvimento da leitura, os princípios linguísticos orientam a seleção dos estímulos, e os métodos psicométricos permitem mensurar, de forma válida e confiável, as habilidades leitoras dos estudantes.

2.2 Psicometria

A psicometria é a área do conhecimento responsável pelo desenvolvimento de métodos e técnicas voltados à mensuração de fenômenos psicológicos e educacionais não diretamente observáveis, denominados construtos ou traços latentes. No contexto educacional, a psicometria é amplamente utilizada para mensurar habilidades, competências, conheci-

mentos e proficiências, fornecendo suporte teórico e estatístico à construção e à validação de instrumentos avaliativos (PASQUALI, 2003).

O desenvolvimento da psicometria está associado à necessidade de produzir medidas mais objetivas e confiáveis do comportamento humano. Segundo Pasquali (2003), a principal finalidade da psicometria consiste em transformar fenômenos subjetivos em variáveis passíveis de análise quantitativa, permitindo a interpretação de características cognitivas e comportamentais por meio de modelos matemáticos e estatísticos.

Na área educacional, a psicometria passou a assumir um papel fundamental diante da expansão das avaliações em larga escala e da necessidade de desenvolver instrumentos capazes de produzir medidas comparáveis entre diferentes populações e contextos. Conforme destacado por Andrade, Tavares e Valle (2000), a utilização de métodos psicométricos possibilita analisar o desempenho dos estudantes de maneira mais precisa, reduzindo a influência de erros de medida e aumentando a confiabilidade dos resultados obtidos.

Um dos modelos psicométricos de maior destaque atualmente é a Teoria de Resposta ao Item (TRI), amplamente utilizada em avaliações educacionais em larga escala devido à sua capacidade de estimar proficiências de forma mais precisa e comparável. No Brasil, a TRI é empregada em exames como o Exame Nacional do Ensino Médio (ENEM), cuja escala de proficiência, proposta por Ferreira (2009), passou a ser calculada com base nesse modelo a partir da reformulação da avaliação em 2009, consolidando-se posteriormente como padrão metodológico para o cálculo das notas do exame.

A Teoria de Resposta ao Item é definida por Andrade, Tavares e Valle (2000) como *“um conjunto de modelos matemáticos que procuram representar a probabilidade de um indivíduo dar uma certa resposta a um item como função dos parâmetros do item e da habilidade (ou das habilidades) do respondente”*. Também se define que alunos com maior habilidade têm maior probabilidade de acertar um item.

Na Teoria de Resposta ao Item, denomina-se “item” cada unidade básica de um instrumento avaliativo utilizada para mensurar determinada habilidade ou traço latente do participante. Em avaliações educacionais, os itens normalmente correspondem a questões, tarefas ou estímulos apresentados aos indivíduos avaliados (ANDRADE; TA-

VARES; VALLE, 2000).

Além disso, de acordo com Pasquali e Primi (2003), a TRI baseia-se na teoria do traço latente, segundo a qual os comportamentos observáveis dos indivíduos são consequência de variáveis hipotéticas não observáveis, denominadas traços latentes ou habilidades. Nesse contexto, o desempenho apresentado nos itens do teste é entendido como uma manifestação dessas habilidades, que podem ser estimadas a partir do padrão de respostas fornecido pelos participantes, em função do nível que o indivíduo possui nesse traço latente.

Ainda de acordo com Andrade, Tavares e Valle (2000), a TRI possui três modelos logísticos para itens dicotômicos, sendo conhecidos como modelos logísticos de 1, 2 e 3 parâmetros, em que cada um considera:

- A dificuldade do item;
- A dificuldade do item e a discriminação;
- A dificuldade do item, a discriminação e a possibilidade de acerto por acaso entre indivíduos de menor habilidade.

2.2.1 Modelo Logístico Unidimensional de 1 Parâmetro

Entre os modelos logísticos apresentados, destaca-se o Modelo Logístico de 1 Parâmetro (1PL), também conhecido como Modelo de Rasch, proposto por Georg Rasch. Nesse modelo, assume-se que todos os itens possuem o mesmo poder de discriminação e que não há influência do acerto ao acaso, de modo que a única característica considerada para cada item é sua dificuldade. Dessa forma, a probabilidade de um indivíduo responder corretamente a um item depende exclusivamente da relação entre sua habilidade e a dificuldade do item.

Segundo Hambleton, Swaminathan e Rogers (1991), o Modelo de Rasch constitui um dos modelos fundamentais da TRI e é amplamente utilizado quando se deseja obter estimativas de habilidade dos indivíduos e de dificuldade dos itens em uma mesma escala métrica, permitindo comparações independentes entre pessoas e itens. Uma de suas principais características é a suposição de invariância dos parâmetros, segundo a qual as

estimativas de habilidade não dependem do conjunto específico de itens utilizados e as estimativas de dificuldade dos itens não dependem da amostra particular de respondentes utilizada na calibração.

De acordo com Andrade, Laros e Gouveia (2010), o desempenho do examinando pode ser modelado por meio de uma função logística crescente denominada Curva Característica do Item (CCI), que relaciona a probabilidade de acerto ao nível do traço latente do indivíduo. No caso do Modelo de Rasch, essa relação é expressa pela Equação 2.1.

$$P(U_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-D(\theta_j - b_i)}}, \quad i = 1, 2, \dots, n \quad (2.1)$$

onde:

- U_{ij} é uma variável dicotômica. Assume 1 se indivíduo j acertar o item i e 0, caso contrário;
- $P(U_{ij} = 1|\theta_j)$ representa a probabilidade de o indivíduo j responder corretamente ao item i ;
- θ_j representa a habilidade do indivíduo j ;
- b_i representa o parâmetro de dificuldade do item i ;
- D é uma constante de escala, geralmente assumida como 1 ou 1,7;
- n é o número de itens do teste.

Observa-se que, quando a habilidade do indivíduo é igual à dificuldade do item ($\theta_j = b_i$), a probabilidade de acerto é de 50%. À medida que a habilidade supera a dificuldade do item, essa probabilidade aumenta, enquanto habilidades inferiores à dificuldade resultam em menores probabilidades de acerto.

2.2.2 Representações Gráficas para Análise Estatística

A análise estatística de dados frequentemente é complementada por representações gráficas, que permitem visualizar padrões, tendências, dispersões e relações entre variáveis de forma

mais intuitiva. Em estudos envolvendo a TRI, essas representações auxiliam tanto na verificação das premissas dos modelos quanto na interpretação dos resultados obtidos.

Uma das representações utilizadas neste trabalho é o *histograma*. Esse gráfico apresenta a distribuição de frequências de uma variável quantitativa por meio da divisão dos dados em intervalos (classes), permitindo identificar características como tendência central, dispersão, assimetria e concentração de observações. No contexto deste estudo, histogramas são empregados para visualizar a distribuição das proficiências estimadas dos estudantes e comparar essa distribuição com a dificuldade dos itens estimada pelo Modelo de Rasch.

Outra representação utilizada é o gráfico de dispersão (*scatter plot*), que consiste na representação simultânea de duas variáveis quantitativas em um plano cartesiano. Esse tipo de gráfico possibilita avaliar visualmente a existência de associação entre variáveis, identificar tendências lineares, agrupamentos e possíveis valores discrepantes (*outliers*). Neste trabalho, gráficos de dispersão são utilizados para comparar a proficiência estimada pela TRI com outras medidas de desempenho, bem como para analisar a relação entre dificuldades observadas e dificuldades previstas por modelos de regressão.

Também é utilizado o gráfico de sedimentação (*scree plot*), amplamente empregado em Análise Fatorial Exploratória para auxiliar na determinação do número de fatores subjacentes aos dados. Nesse gráfico, os autovalores são apresentados em ordem decrescente, permitindo identificar o ponto em que ocorre estabilização da curva, conhecido como “cotovelo” (*elbow*).

Além dessas representações, curvas acumuladas podem ser empregadas para analisar o comportamento de uma variável ao longo de uma sequência ordenada de observações. Diferentemente dos histogramas, essas curvas representam valores acumulados, permitindo comparar a distribuição progressiva de uma característica entre diferentes grupos. Neste trabalho, essa abordagem é utilizada para avaliar a distribuição acumulada da dificuldade dos itens ao longo das posições ocupadas nos diferentes cadernos da avaliação, possibilitando verificar o equilíbrio entre as versões do instrumento.

3 Trabalhos Relacionados

Neste capítulo, são apresentados trabalhos relacionados à aplicação de avaliações de leitura de palavras e à utilização da Teoria de Resposta ao Item (TRI) para análise psicométrica e para o cálculo da dificuldade dos itens. Os estudos selecionados abordam diferentes perspectivas relacionadas à mensuração da habilidade leitora, incluindo investigações sobre discriminação de itens, estimação de proficiência, funcionamento psicométrico de testes de leitura e análise de características linguísticas associadas à dificuldade das palavras.

Além disso, este levantamento busca evidenciar como modelos dicotômicos da TRI vêm sendo empregados na literatura para analisar o desempenho de indivíduos em tarefas de leitura, permitindo identificar as limitações e potencialidades dos instrumentos avaliativos empregados em contextos educacionais. Dessa forma, os trabalhos apresentados contribuem para fundamentar teoricamente a presente pesquisa, bem como para situá-la no contexto das discussões atuais sobre a avaliação da leitura, a modelagem psicométrica e a mensuração de habilidades educacionais.

3.1 Avaliação da Leitura de Palavras e Desenvolvimento de Instrumentos

A leitura de palavras isoladas é amplamente utilizada em instrumentos destinados à avaliação do desenvolvimento da alfabetização, por permitir a mensuração relativamente direta de habilidades relacionadas à decodificação e ao reconhecimento ortográfico. No contexto brasileiro, diversos instrumentos utilizam essa abordagem, incluindo o Teste de Desempenho Escolar (TDE), sua versão atualizada, o TDE-II, e a Prova de Leitura e Escrita de Palavras (PLEP).

Embora compartilhem o objetivo de avaliar a habilidade de leitura, os estudos apresentam enfoques distintos. Enquanto Athayde et al. (2019) concentraram-se na construção e validação do Subteste de Leitura do TDE-II, buscando aperfeiçoar aspectos psi-

cométricos e psicolinguísticos do instrumento, Knijnik et al. (2014) e Lúcio et al. (2018) investigaram o funcionamento dos itens por meio da Teoria de Resposta ao Item, analisando parâmetros como discriminação, dificuldade e informação. Esses trabalhos convergem ao demonstrar a importância da leitura de palavras isoladas como indicador do desenvolvimento das habilidades básicas. Além disso, apontam para a necessidade de instrumentos capazes de discriminar adequadamente estudantes com diferentes níveis de proficiência, combatendo o efeito de teto observado em versões anteriores, causado pela concentração excessiva de itens fáceis

A qualidade psicométrica desses instrumentos está intrinsecamente ligada às características psicolinguísticas das palavras selecionadas. Estudos de TRI evidenciam que a dificuldade dos itens não é aleatória, mas determinada por variáveis como frequência lexical, regularidade ortográfica, extensão da palavra e idade de aquisição (COMPTON et al., 2025; LÚCIO et al., 2018). Dessa forma, a precisão das medidas obtidas depende da distribuição criteriosa desses estímulos ao longo do continuum de habilidade (theta), garantindo que o teste forneça informação máxima tanto para leitores iniciantes quanto para os mais proficientes (KNIJNIK et al., 2014).

3.2 Aplicação da Teoria de Resposta ao Item em Avaliações de Leitura

A Teoria de Resposta ao Item tem sido amplamente utilizada para investigar o comportamento psicométrico de instrumentos de leitura e para estimar parâmetros associados às palavras avaliadas. Nos estudos brasileiros, a TRI tem sido empregada principalmente como ferramenta de validação e aperfeiçoamento de instrumentos já existentes.

Knijnik et al. (2014), por exemplo, utilizaram modelos da TRI para analisar os subtestes de leitura e escrita do TDE, identificando quais regiões do continuum de habilidade eram adequadamente mensuradas pelos itens. De forma semelhante, Lúcio et al. (2018) aplicaram modelos de TRI a uma tarefa de decodificação composta por palavras de baixa frequência, investigando a dificuldade e a discriminação dos itens e propondo uma versão reduzida do instrumento baseada nos resultados psicométricos obtidos.

Embora esses trabalhos demonstrem o potencial da TRI para avaliação da qualidade dos instrumentos, seu foco principal permanece na análise do teste como um todo. Em contraste, estudos mais recentes passaram a direcionar atenção para a própria palavra como unidade de análise. Um exemplo dessa abordagem é o Developmental English Lexicon Project (d-ELP), que construiu uma base contendo estimativas contínuas de dificuldade para aproximadamente dez mil palavras da língua inglesa, obtidas por meio de modelos Rasch aplicados a respostas de milhares de estudantes (COMPTON et al., 2025). Além da estimação da dificuldade, o projeto buscou relacionar características linguísticas das palavras ao desempenho observado, permitindo investigar fatores associados à complexidade da leitura em nível lexical.

Essa diferença de enfoque evidencia uma importante distinção entre os estudos encontrados na literatura. Enquanto os trabalhos brasileiros concentram-se predominantemente na avaliação psicométrica de instrumentos, iniciativas como o d-ELP exploram a dificuldade das palavras como fenômeno de interesse próprio, investigando quais características linguísticas e cognitivas modulam o processamento lexical ao longo do desenvolvimento da fluência leitora.

3.3 Características Linguísticas Associadas à Dificuldade de Leitura

Compton et al. (2025) apontam que a dificuldade de leitura não depende exclusivamente da habilidade do estudante, sendo também determinada pelas características intrínsecas das palavras avaliadas. Consequentemente, a seleção de itens para avaliações de fluência requer o controle rigoroso de variáveis relacionadas ao processamento linguístico. No desenvolvimento do Subteste de Leitura do TDE-II, foram considerados critérios como extensão lexical, estrutura silábica, grau de familiaridade e o nível de aprendizagem esperado para cada ano escolar (ATHAYDE et al., 2019).

A literatura destaca a relevância da estrutura silábica no desenvolvimento leitor, indicando que palavras com estruturas simples tendem a ser reconhecidas com maior facilidade, enquanto padrões não canônicos e encontros consonantais aumentam a carga

de processamento. No contexto do português brasileiro, a predominância de sílabas do tipo Consoante-Vogal (CV) favorece a decodificação inicial, especialmente quando comparada a idiomas com profundidade ortográfica maior e estruturas silábicas mais complexas (SEYMOUR; ARO; ERSKINE, 2003). Além da estrutura, a frequência lexical é uma variável determinante; o uso de palavras de baixa frequência, como proposto na PLEP, é uma estratégia eficaz para evitar o efeito de teto e discriminar melhor leitores com diferentes níveis de proficiência (LÚCIO et al., 2018).

Aprofundando a análise das variáveis psicolinguísticas, o projeto Developmental English Lexicon Project (d-ELP) estabeleceu estimativas contínuas de dificuldade para milhares de palavras utilizando o modelo Rasch da TRI (COMPTON et al., 2025). O projeto investigou o impacto de preditores como frequência, comprimento da palavra, vizinhança ortográfica, semântica e transparência ortográfica (consistência entre grafema e fonema) no desempenho leitor. Os resultados do d-ELP demonstraram que, embora múltiplos fatores contribuam para a complexidade lexical, a Idade de Aquisição (AoA) e a transparência ortográfica são os preditores que explicam a maior parte da variância única na dificuldade de leitura.

Observa-se, portanto, que as características das palavras desempenham um papel central na determinação da dificuldade dos itens e na precisão dos instrumentos. No entanto, apesar da robustez das evidências internacionais, o cenário nacional ainda carece de investigações que relacionem sistematicamente essas variáveis psicolinguísticas aos parâmetros de dificuldade estimados por modelos psicométricos em larga escala, como constatado por Lúcio et al. (2018).

3.4 Análise Comparativa

A Tabela 3.1 apresenta um comparativo entre os trabalhos relacionados discutidos anteriormente, considerando aspectos relevantes para o desenvolvimento desta pesquisa, como o uso da Teoria de Resposta ao Item (TRI), a análise de dificuldade lexical, a utilização de leitura de palavras isoladas e a investigação de características linguísticas dos itens.

Como pode ser observado na Tabela 3.1, todos os trabalhos analisados utilizam a Teoria de Resposta ao Item como principal abordagem psicométrica para análise de

Tabela 3.1: Tabela comparativa entre trabalhos relacionados

Trabalho	TRI	Palavras isoladas	Dificuldade	Caract. linguísticas
<i>Avaliação dos subtestes de leitura e escrita do teste de desempenho escolar através da Teoria de Resposta ao Item</i> (Knijnik et al., 2014)	X	X	X	
<i>Desenvolvimento do subteste de leitura do Teste de Desempenho Escolar II</i> (Athayde et al., 2019)	X	X	X	X
<i>Word decoding task: item analysis by IRT and within-group norms</i> (Lúcio et al., 2019)	X	X	X	
<i>The Developmental English Lexicon Project (d-ELP): Continuous Word Reading Difficulty Estimates of the 9,961 Most Frequently Printed English Words for US Children</i> (Compton et al., 2025)	X	X	X	X

desempenho em leitura. Além disso, todos os estudos trabalham diretamente com a leitura de palavras isoladas, o que evidencia a relevância desse tipo de tarefa para a avaliação da habilidade leitora.

Também é possível observar que todos os trabalhos realizam algum tipo de análise da dificuldade dos itens, seja para a validação psicométrica do instrumento, seja para a estimativa contínua da dificuldade lexical das palavras. Entretanto, apenas os trabalhos de Athayde et al. (2019) e Compton et al. (2025) exploram explicitamente características linguísticas das palavras, como frequência, regularidade ortográfica e propriedades fonológicas, buscando compreender fatores associados à dificuldade de leitura.

Dessa forma, nota-se que ainda há poucas pesquisas que relacionam simultaneamente TRI, a leitura de palavras isoladas e a análise detalhada das características linguísticas dos itens. Além disso, nenhum dos trabalhos analisados investiga diretamente possíveis efeitos da posição das palavras no teste sobre o desempenho dos participantes, o que constitui um dos diferenciais da presente pesquisa.

3.5 Considerações Finais

Os trabalhos analisados demonstram que a Teoria de Resposta ao Item vem sendo amplamente utilizada em pesquisas relacionadas à avaliação educacional e, mais especificamente, à avaliação da leitura de palavras isoladas. Os estudos apresentados evidenciam que a TRI possibilita estimativas mais precisas da habilidade dos participantes e da dificuldade dos itens, permitindo análises psicométricas mais robustas do que as obtidas exclusivamente pela Teoria Clássica dos Testes.

Além disso, observa-se que parte dos trabalhos recentes busca relacionar parâmetros psicométricos a características linguísticas das palavras, como a frequência lexical, a regularidade ortográfica e as propriedades fonológicas. Essa abordagem contribui para uma compreensão mais aprofundada dos fatores associados à dificuldade de leitura e ao desenvolvimento da habilidade de leitura.

Apesar dos avanços identificados na literatura, ainda existem lacunas na investigação dos fatores contextuais envolvidos na aplicação dos testes, especialmente quanto ao possível efeito da posição dos itens sobre o desempenho dos estudantes. Também são limitados os estudos que buscam integrar simultaneamente a análise psicométrica, a leitura de palavras isoladas e a investigação detalhada de habilidades linguísticas específicas.

Dessa forma, a presente pesquisa busca contribuir para a área ao propor uma análise baseada na TRI aplicada à leitura de palavras isoladas, considerando não apenas a dificuldade dos itens e a habilidade dos participantes, mas também as possíveis influências da posição das palavras no instrumento e a relação entre características linguísticas e desempenho leitor. Espera-se, assim, fornecer evidências que auxiliem no desenvolvimento de instrumentos de avaliação mais precisos, interpretáveis e adequados ao contexto educacional.

4 Metodologia

Este capítulo apresenta os procedimentos metodológicos adotados para a construção da base analítica utilizada neste trabalho, bem como as etapas de tratamento, limpeza, reorganização e modelagem dos dados. A metodologia foi organizada com base no uso de dados secundários provenientes de uma avaliação educacional de fluência em leitura, com foco na análise das respostas dos estudantes a palavras isoladas, por meio da Teoria de Resposta ao Item (TRI).

4.1 Caracterização da pesquisa

A presente pesquisa adota uma abordagem quantitativa, uma vez que utiliza registros estruturados de respostas de estudantes para realizar análises estatísticas e psicométricas. Além disso, caracteriza-se como uma pesquisa aplicada, pois busca investigar um problema relacionado à avaliação educacional, mais especificamente à estimação da proficiência em leitura de palavras isoladas e à análise da dificuldade dos itens que compõem o instrumento avaliativo.

O estudo utiliza dados secundários provenientes das avaliações PARC Entrada 2024 e PARC Saída 2024, cuja estrutura inclui registros de correção de áudios de leitura. A partir desses dados, foram realizados procedimentos de limpeza, transformação e padronização com o objetivo de construir uma matriz de respostas adequada à aplicação de modelos da Teoria de Resposta ao Item.

4.2 Caracterização da avaliação de leitura de palavras

A avaliação analisada neste trabalho tem como objetivo mensurar o desempenho dos estudantes em uma tarefa de leitura oral de palavras isoladas. Nesse tipo de instrumento, o estudante é apresentado a um conjunto de palavras e deve realizar a leitura em voz

alta. A produção oral é registrada em áudio e, posteriormente, submetida a um processo de correção, no qual cada palavra lida recebe uma classificação com base no desempenho observado.

No caso da avaliação considerada, cada estudante responde a um dos cadernos disponíveis para aplicação. Esses cadernos contêm um conjunto de 60 palavras, organizadas em diferentes posições conforme a versão do instrumento. Assim, embora as palavras avaliadas sejam as mesmas, a ordem em que aparecem pode variar entre os cadernos. Essa característica é relevante para este trabalho, pois permite investigar, em etapas posteriores, se a posição ocupada pela palavra no caderno pode influenciar a probabilidade de acerto do estudante.

A correção das leituras pode ser realizada automaticamente, por meio de modelos de inteligência artificial, ou manualmente, por corretores humanos. Cada palavra recebe um código de correção, indicando a classificação atribuída à leitura realizada pelo estudante. Entre as categorias possíveis, estão situações como acerto, silabação da palavra, nomeação de letras, invenção de palavra, erro e ausência de leitura.

Como diferentes projetos podem adotar fluxos distintos de correção, os registros podem apresentar tanto correção automática quanto correção humana. Neste trabalho, a identificação do tipo de correção foi utilizada principalmente para o tratamento de duplicidades, com o objetivo de manter uma única correção por estudante e por caderno.

Dessa forma, a avaliação de leitura de palavras consiste em uma estrutura composta por estudantes, cadernos, palavras e códigos de correção. A partir dessa estrutura, foi construída uma base analítica em que cada linha representa um estudante e cada coluna representa uma palavra avaliada, o que possibilita a aplicação de modelos da Teoria de Resposta ao Item.

4.3 Descrição das bases de dados

Os dados utilizados neste trabalho são provenientes de uma avaliação de fluência leitora aplicada a estudantes do **2º ano do Ensino Fundamental**, no âmbito dos projetos **PARC Entrada 2024** e **PARC Saída 2024**. O instrumento teve como objetivo avaliar a capacidade de leitura oral de palavras isoladas, permitindo a obtenção de informações

sobre o desempenho dos estudantes em habilidades relacionadas ao reconhecimento e à decodificação de palavras.

Cabe destacar que as avaliações foram aplicadas em dois momentos distintos do ano letivo de 2024: no início e no final do período escolar. Espera-se que um mesmo estudante participe de ambas as aplicações, permitindo a comparação de seu desempenho ao ingressar e ao concluir a série avaliada. Dessa forma, torna-se possível acompanhar a evolução das habilidades de leitura ao longo do ano letivo.

Cada avaliação foi composta por **60 palavras** distribuídas em **5 cadernos** distintos. Embora todos os cadernos contivessem o mesmo conjunto de palavras, a ordem de apresentação dos itens variava entre as versões do instrumento, minimizando possíveis efeitos associados à posição das palavras durante a aplicação da avaliação.

Inicialmente, a base de dados era composta por **2.904.548 registros** de respostas. Após os procedimentos de limpeza, consistência e tratamento de duplicidades, descritos nas seções seguintes, a base final passou a conter **2.557.162 registros válidos**, cada um correspondendo a um estudante.

Cada estudante realizou a leitura oral das palavras presentes no caderno recebido, sendo as respostas registradas e posteriormente corrigidas. Os resultados dessas correções constituíram a base utilizada para a construção da matriz de respostas empregada nas análises psicométricas realizadas neste trabalho.

A Tabela 4.1 apresenta um resumo das principais características da avaliação utilizada.

Tabela 4.1: Caracterização das avaliações analisadas

Característica	Valor
Projetos analisados	PARC Entrada 2024 e PARC Saída 2024
Ano escolar	2º ano do Ensino Fundamental
Quantidade de palavras avaliadas	60 palavras
Quantidade de cadernos	5 cadernos
Registros após limpeza (Entrada 2024)	1.269.664 registros
Registros após limpeza (Saída 2024)	1.287.498 registros

Fonte: Elaborado pelo autor.

4.4 Codificação das respostas

Para a aplicação do modelo psicométrico, as respostas foram convertidas para uma escala dicotômica. Nessa etapa, as respostas classificadas como corretas receberam valor 1, enquanto as respostas incorretas receberam valor 0. As respostas correspondentes a palavras não lidas foram tratadas como dados ausentes (*missing values*), uma vez que, devido à natureza do instrumento, não era esperado que todos os estudantes realizassem a leitura de todas as palavras apresentadas. Dessa forma, essas observações foram interpretadas como itens não respondidos, sendo representadas pelo valor NA.

A Tabela 4.2 apresenta a regra utilizada nessa etapa.

Tabela 4.2: Codificação preliminar das respostas

Código original	Interpretação	Valor
1	Acertou	1
2	Silabou a palavra	1
3	Nomeou letras da palavra	0
4	Inventou palavra	0
5	Errou	0
NÃO LEU	Palavra não lida	NA

Fonte: Elaborado pelo autor.

4.5 Aplicação da Teoria de Resposta ao Item

Após os procedimentos de preparação e tratamento dos dados, foi construída uma matriz de respostas dicotômicas para cada projeto, em que cada linha representa um estudante e cada coluna representa uma palavra avaliada.

Antes da aplicação da Teoria de Resposta ao Item, foi avaliada a adequação da hipótese de unidimensionalidade, pressuposto fundamental do Modelo de Rasch. Segundo Hambleton, Swaminathan e Rogers (1991), espera-se que exista um fator dominante, indicando que a maior parte da variabilidade das respostas pode ser atribuída a um único construto latente.

Como os itens são dicotômicos, a análise foi realizada por meio de Análise Fatorial

Exploratória (AFE), utilizando uma matriz de correlações tetracóricas, abordagem mais apropriada para variáveis desse tipo. Inicialmente, foram obtidos os autovalores da matriz de correlação e analisado o gráfico de sedimentação (*scree plot*), com o objetivo de verificar a existência de um fator predominante. Em seguida, foi estimada uma solução unifatorial para inspeção das cargas fatoriais dos itens.

Adicionalmente, foi calculada a razão entre o primeiro e o segundo autovalor como medida descritiva da predominância do primeiro fator. Essa etapa teve como objetivo verificar se a maior parte da variabilidade das respostas poderia ser atribuída a um único construto latente relacionado à habilidade de leitura de palavras isoladas.

Após a verificação da unidimensionalidade, a estimação dos parâmetros foi realizada na linguagem R utilizando o pacote **TAM**, amplamente empregado para análises psicométricas baseadas na Teoria de Resposta ao Item. A partir da matriz de respostas, foram estimados simultaneamente os parâmetros de dificuldade dos itens e as proficiências dos estudantes.

Neste trabalho foi adotado o Modelo Logístico de Um Parâmetro (1PL), também conhecido como Modelo de Rasch. A escolha desse modelo foi motivada pelas características da avaliação analisada, composta por itens dicotômicos de leitura oral de palavras isoladas. Nesse contexto, a ocorrência de acertos ao acaso é considerada pouco provável, uma vez que os estudantes não selecionam alternativas, mas devem realizar a leitura da palavra apresentada. Além disso, o modelo assume discriminação constante entre os itens, permitindo que a dificuldade das palavras seja o principal parâmetro analisado.

A partir da aplicação do modelo, foram estimados os parâmetros de dificuldade (b_i) para cada uma das palavras avaliadas e os parâmetros de proficiência (θ_j) para os estudantes participantes da avaliação. As estimativas de dificuldade permitiram posicionar as palavras em uma escala comum, possibilitando identificar quais itens tendem a apresentar maior ou menor probabilidade de acerto para estudantes com diferentes níveis de habilidade.

Para as análises inferenciais realizadas neste trabalho, adotou-se nível de significância de 5% ($\alpha = 0,05$). Assim, resultados com valor-p inferior a 0,05 foram con-

siderados estatisticamente significativos, indicando evidências suficientes para rejeição da hipótese nula. Valores-p compreendidos entre 0,05 e 0,10 foram interpretados como tendência estatística, indicando associações que, embora não atendam ao critério convencional de significância, sugerem possíveis efeitos que merecem investigação em estudos futuros. Os valores-p apresentados nos modelos de regressão linear foram obtidos por meio do teste t aplicado aos coeficientes estimados.

As estimativas obtidas por meio da TRI constituem a base para as análises posteriores desenvolvidas neste trabalho. Em particular, os parâmetros de dificuldade foram utilizados para investigar possíveis relações entre características linguísticas das palavras e sua dificuldade de leitura, bem como para analisar a distribuição de dificuldade dos itens nos diferentes cadernos de avaliação. De forma complementar, as estimativas de proficiência dos estudantes foram utilizadas para examinar a distribuição das habilidades de leitura e sua relação com indicadores tradicionais de desempenho, como a quantidade de palavras lidas corretamente.

5 Análise dos resultados

Neste capítulo são apresentados os resultados obtidos a partir da aplicação da Teoria de Resposta ao Item aos dados de leitura de palavras isoladas. Inicialmente, são analisados os parâmetros de dificuldade estimados para as palavras avaliadas, buscando compreender como características linguísticas dos itens podem estar associadas aos diferentes níveis de dificuldade observados. Em seguida, é investigada a distribuição da dificuldade ao longo dos cadernos de avaliação, considerando que as mesmas palavras são apresentadas em diferentes ordens e que essa organização pode influenciar a experiência do estudante durante a aplicação do teste. Por fim, são examinadas as distribuições de proficiência estimadas para os estudantes e sua relação com indicadores tradicionais de desempenho, permitindo comparar as estimativas produzidas pela TRI com medidas baseadas na quantidade de palavras lidas corretamente. Os resultados apresentados contribuem para a compreensão das características dos itens e dos participantes avaliados, além de fornecer evidências sobre a adequação do modelo psicométrico adotado neste trabalho.

5.1 Verificação da Unidimensionalidade

Conforme discutido por Hambleton, Swaminathan e Rogers (1991), a hipótese de unidimensionalidade não exige que apenas um fator influencie o desempenho dos estudantes. Na prática, fatores como motivação, ansiedade, velocidade de resposta e outras habilidades cognitivas também podem afetar as respostas aos itens. O pressuposto do Modelo de Rasch é considerado adequadamente atendido quando existe um fator dominante responsável pela maior parte da variabilidade observada nas respostas. Dessa forma, a análise realizada neste trabalho buscou verificar a predominância de um único fator latente associado à habilidade de leitura de palavras isoladas.

Como os itens são dicotômicos, assumindo os valores 0 (erro) e 1 (acerto), e podem apresentar respostas ausentes decorrentes do desenho da avaliação, a análise foi conduzida utilizando uma matriz de correlações tetracóricas, mais apropriada para esse

tipo de variável do que a matriz de correlações de Pearson.

Inicialmente, foi realizada uma análise dos autovalores da matriz de correlação, acompanhada da inspeção do gráfico de sedimentação (*scree plot*). Também foi calculada a razão entre o primeiro e o segundo autovalor, adotando-se como referência o critério de que o primeiro autovalor deve ser, no mínimo, quatro vezes superior ao segundo para indicar predominância de um fator geral. Complementarmente, foi estimada uma solução unifatorial para inspeção das cargas fatoriais dos itens.

As Tabelas 5.1 e 5.2 apresentam os principais resultados obtidos para as avaliações PARC Entrada 2024 e PARC Saída 2024, respectivamente.

Tabela 5.1: Resultados da análise de dimensionalidade – PARC Entrada 2024

Indicador	Valor
Primeiro autovalor	37,01
Segundo autovalor	4,25
Terceiro autovalor	1,33
Razão λ_1/λ_2	8,72
Variância explicada pelo primeiro fator	61,69%
Variância explicada pelo segundo fator	7,08%
Variância explicada pelo terceiro fator	2,22%
Menor carga fatorial	0,38
Maior carga fatorial	0,93

Fonte: Elaborado pelo autor.

Tabela 5.2: Resultados da análise de dimensionalidade – PARC Saída 2024

Indicador	Valor
Primeiro autovalor	39,47
Segundo autovalor	3,10
Terceiro autovalor	1,03
Razão λ_1/λ_2	12,75
Variância explicada pelo primeiro fator	65,79%
Variância explicada pelo segundo fator	5,16%
Variância explicada pelo terceiro fator	1,71%
Menor carga fatorial	0,54
Maior carga fatorial	0,94

Fonte: Elaborado pelo autor.

A Figura 5.1 apresenta os gráficos de sedimentação obtidos para as duas avaliações.

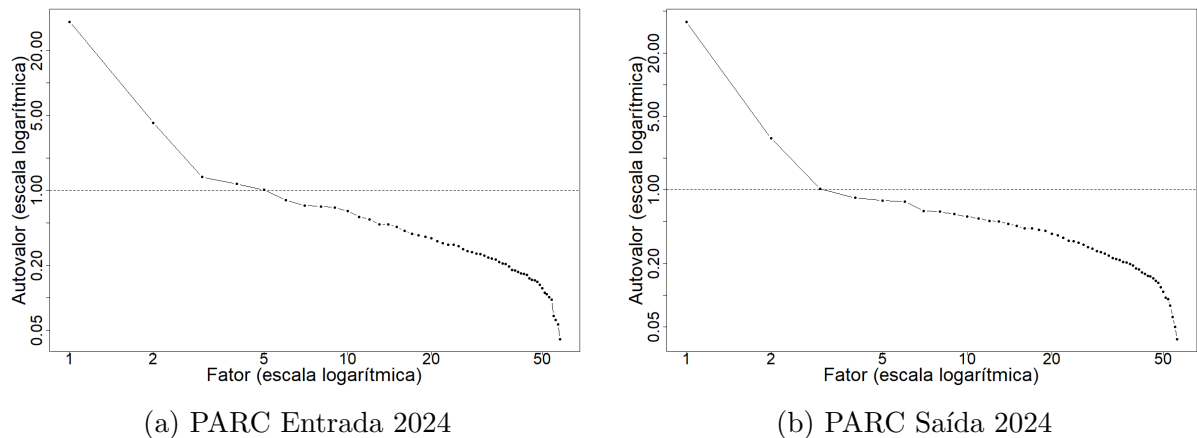


Figura 5.1: Gráficos de sedimentação.

Fonte: Elaborado pelo autor.

Observa-se, em ambas as avaliações, um acentuado predomínio do primeiro fator em relação aos demais. Na PARC Entrada 2024, o primeiro autovalor foi aproximadamente 8,7 vezes superior ao segundo, enquanto na PARC Saída 2024 essa razão atingiu aproximadamente 12,7. Em ambos os casos, a razão entre os dois primeiros autovalores superou com ampla margem o critério de quatro vezes, indicando predominância de um

único fator latente.

Além disso, o primeiro fator explicou 61,69% da variância total na PARC Entrada 2024 e 65,79% na PARC Saída 2024, valores substancialmente superiores aos observados para os fatores subsequentes. Nos gráficos de sedimentação, observa-se uma acentuada descontinuidade entre o primeiro e o segundo autovalor, seguida por uma estabilização gradual da curva. Esse comportamento caracteriza a presença de um fator predominante, indicando que grande parte da covariância observada entre os itens pode ser explicada por um único construto latente.

A inspeção da solução unifatorial também apresentou resultados consistentes com a hipótese de unidimensionalidade. Na PARC Entrada 2024, as cargas fatoriais variaram entre 0,38 e 0,93, enquanto na PARC Saída 2024 variaram entre 0,54 e 0,94. Em ambas as avaliações, todos os itens apresentaram cargas fatoriais moderadas ou elevadas, indicando associação consistente com o fator geral de leitura.

Em conjunto, esses resultados fornecem evidências de que o conjunto de itens atende de forma satisfatória ao pressuposto de unidimensionalidade requerido para a aplicação do Modelo de Rasch, justificando a utilização das estimativas de dificuldade e proficiência apresentadas nas seções seguintes.

Além da unidimensionalidade, a Teoria de Resposta ao Item pressupõe independência local entre os itens. Segundo Hambleton, Swaminathan e Rogers (1991), quando um fator dominante explica adequadamente o desempenho dos estudantes, a independência local tende a ser aproximadamente satisfeita. Assim, as evidências favoráveis de unidimensionalidade observadas neste trabalho também fornecem suporte indireto para esse pressuposto do Modelo de Rasch.

5.2 Distribuição da Dificuldade dos Cadernos

Embora os cinco cadernos analisados contenham exatamente o mesmo conjunto de palavras, a ordem de apresentação dos itens varia entre as diferentes versões da avaliação. Dessa forma, apesar de a dificuldade total dos cadernos ser teoricamente equivalente, a distribuição dessa dificuldade ao longo da aplicação pode diferir significativamente. Essa característica é particularmente relevante em avaliações de fluência leitora, uma vez que

fatores como fadiga, tempo disponível para resposta e progressão da complexidade dos itens podem influenciar o desempenho dos estudantes.

Com o objetivo de investigar como a dificuldade das palavras está distribuída em cada versão da avaliação, foram utilizadas as estimativas de dificuldade obtidas por meio da Teoria de Resposta ao Item. A partir dessas estimativas, analisou-se a progressão da dificuldade ao longo das posições de cada caderno, permitindo identificar possíveis diferenças na forma como itens mais fáceis e mais difíceis foram organizados. Essa análise possibilita verificar se os cadernos apresentam distribuições semelhantes de dificuldade ou se determinadas versões concentram palavras mais difíceis em regiões específicas da avaliação.

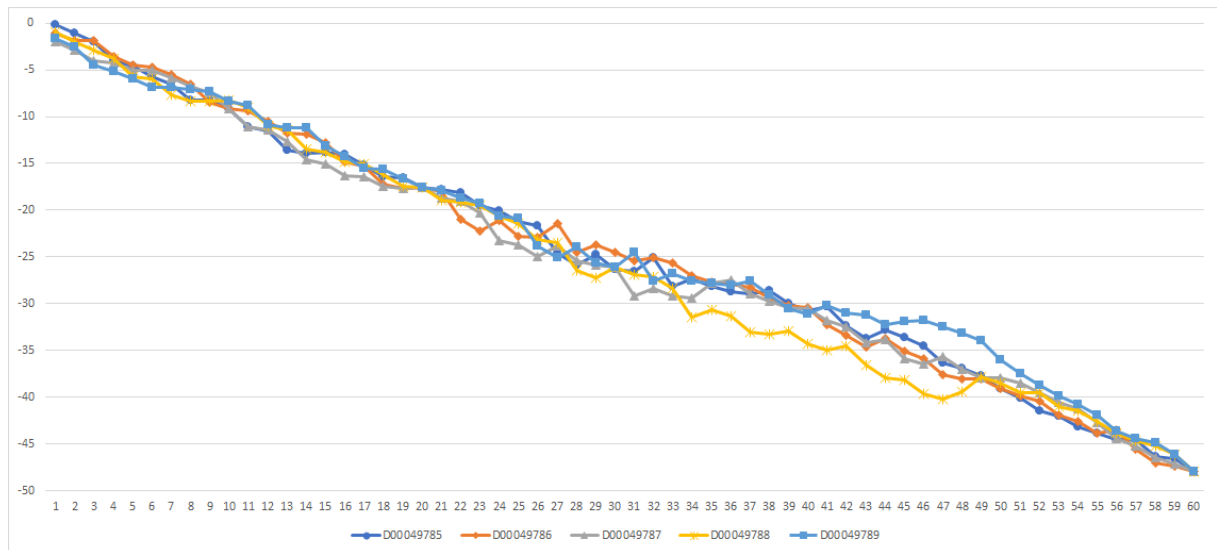
5.2.1 Distribuição Acumulada da Dificuldade dos Cadernos

Com o objetivo de analisar como a dificuldade dos itens está distribuída ao longo dos diferentes cadernos da avaliação, foram calculadas as dificuldades acumuladas das palavras em função de sua posição de apresentação. Embora todos os cadernos contenham o mesmo conjunto de palavras, a ordem de apresentação dos itens varia entre as versões, o que pode resultar em diferentes progressões de dificuldade ao longo da aplicação.

Para cada caderno, as palavras foram inicialmente ordenadas conforme a posição em que aparecem no instrumento. Em seguida, calculou-se a soma acumulada dos parâmetros de dificuldade (b) estimados pela TRI ao longo das 60 posições. Dessa forma, cada ponto da curva representa a dificuldade total acumulada até determinada posição do caderno. Caso os diferentes cadernos apresentem curvas semelhantes, isso indica que a distribuição de palavras fáceis e difíceis permanece equilibrada ao longo da avaliação, ainda que a ordem de apresentação varie entre eles.

A Figura 5.2 apresenta as curvas de dificuldade acumulada para os cadernos da avaliação PARC Entrada 2024. Observa-se que as curvas apresentam comportamento bastante semelhante, indicando que a distribuição da dificuldade foi realizada de forma relativamente equilibrada entre os diferentes cadernos. Pequenas divergências podem ser observadas em posições intermediárias da avaliação, porém essas diferenças tendem a se reduzir ao longo do teste.

Figura 5.2: Distribuição acumulada da dificuldade dos cadernos da avaliação PARC Entrada 2024

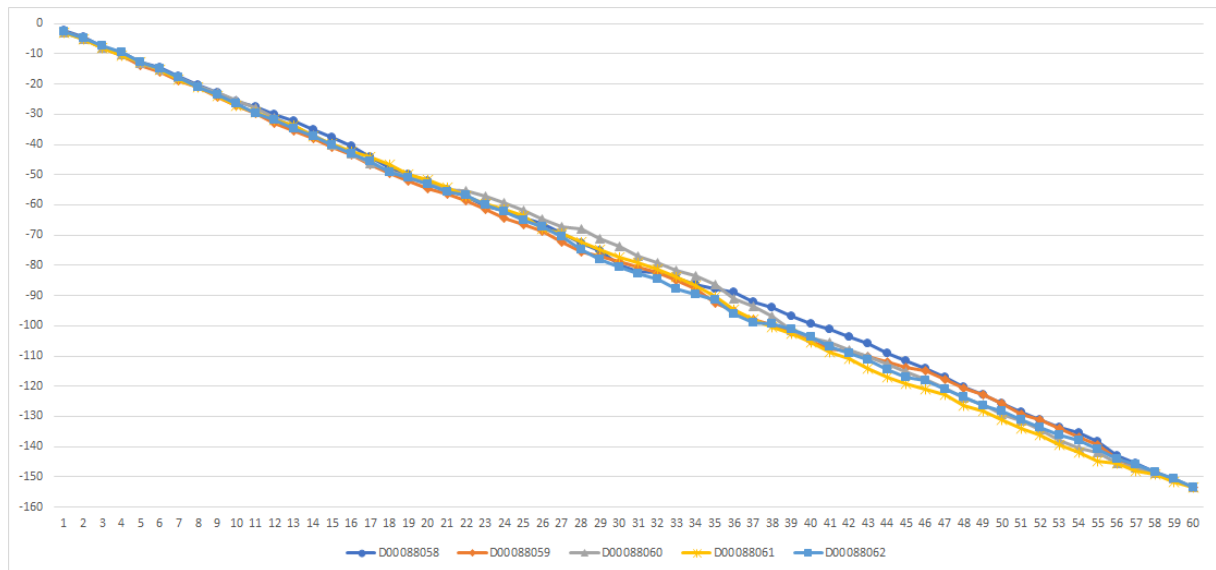


Fonte: Elaborado pelo autor.

Esse comportamento era esperado, uma vez que todos os cadernos são compostos pelas mesmas palavras. Dessa forma, embora a sequência de apresentação dos itens seja diferente, a dificuldade total acumulada converge para valores semelhantes ao final da avaliação.

Resultado semelhante é observado na Figura 5.3, referente à avaliação PARC Saída 2024. As curvas apresentam elevada proximidade ao longo de praticamente toda a extensão da avaliação, sugerindo que as diferentes versões do instrumento distribuem a dificuldade dos itens de maneira comparável. As maiores diferenças ocorrem apenas em trechos específicos do teste e não indicam a existência de um caderno consistentemente mais difícil ou mais fácil que os demais.

Figura 5.3: Distribuição acumulada da dificuldade dos cadernos da avaliação PARC Saída 2024



Fonte: Elaborado pelo autor.

5.2.2 Dificuldade dos Itens por Posição

Além da análise acumulada, foi investigada a dificuldade dos itens em cada posição dos cadernos. Essa análise permite identificar variações locais na distribuição dos itens e compreender em quais momentos da avaliação ocorrem as maiores diferenças entre as versões do instrumento.

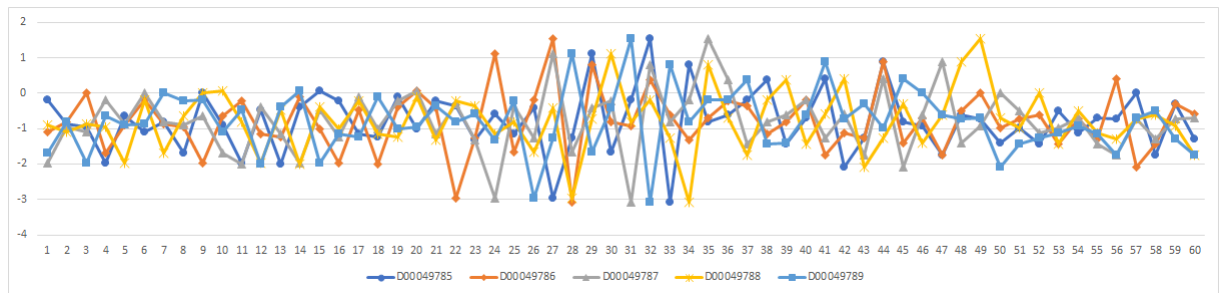
A Figura 5.4 apresenta a dificuldade das palavras em cada posição dos cadernos da avaliação PARC Entrada 2024. Verifica-se que a maior parte das palavras apresenta níveis de dificuldade relativamente próximos entre os diferentes cadernos. Entretanto, algumas posições específicas exibem maior dispersão, indicando que determinadas versões receberam palavras relativamente mais difíceis ou mais fáceis nesses pontos da avaliação.

Apesar dessas oscilações pontuais, não se observa um padrão consistente que caracterize um dos cadernos como sistematicamente mais difícil que os demais. As diferenças encontradas parecem decorrer principalmente da redistribuição das palavras ao longo das posições e não de uma concentração persistente de itens difíceis em um único caderno.

A mesma tendência é observada na avaliação PARC Saída 2024, na Figura 5.5. As dificuldades estimadas apresentam oscilações ao longo das posições, porém as curvas

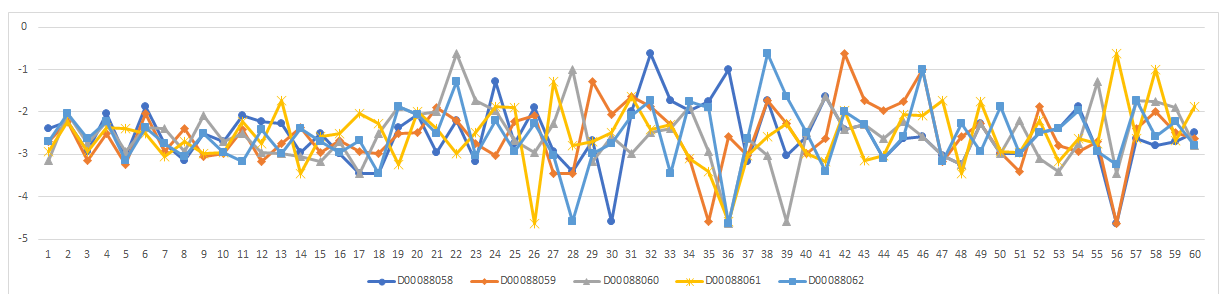
permanecem próximas entre si durante toda a extensão do teste. Esse resultado sugere que a estratégia de montagem dos cadernos conseguiu manter uma distribuição relativamente homogênea da dificuldade dos itens, mesmo com a alteração da ordem de apresentação das palavras.

Figura 5.4: Distribuição da dificuldade dos itens por posição nos cadernos da avaliação PARC Entrada 2024



Fonte: Elaborado pelo autor.

Figura 5.5: Distribuição da dificuldade dos itens por posição nos cadernos da avaliação PARC Saída 2024



Fonte: Elaborado pelo autor.

5.3 Distribuição da Proficiência dos Estudantes

Com o objetivo de investigar a distribuição das habilidades de leitura dos estudantes e sua relação com medidas tradicionais de desempenho, foram analisadas as proficiências estimadas pela Teoria de Resposta ao Item e a quantidade de palavras corretas (QPC) observada nas avaliações PARC Entrada 2024 e PARC Saída 2024.

A análise foi conduzida em duas etapas. Inicialmente, foram examinadas as distribuições de proficiência e de QPC em cada aplicação da avaliação. Em seguida, foi investigada a relação entre essas duas medidas por meio de análises de correlação, do ajuste

de modelos de regressão linear simples estimados pelo método dos Mínimos Quadrados Ordinários (*Ordinary Least Squares* - OLS) e de visualizações gráficas. Dessa forma, buscou-se verificar em que medida a proficiência estimada pela TRI é consistente com a métrica tradicional baseada na contagem de acertos e como essa relação se manifesta nos diferentes momentos da avaliação.

5.3.1 Comparação entre PARC Entrada 2024 e PARC Saída 2024

A Figura 5.6 apresenta as distribuições das proficiências estimadas pelo modelo de Rasch para os estudantes participantes das avaliações PARC Entrada 2024 e PARC Saída 2024, respectivamente.

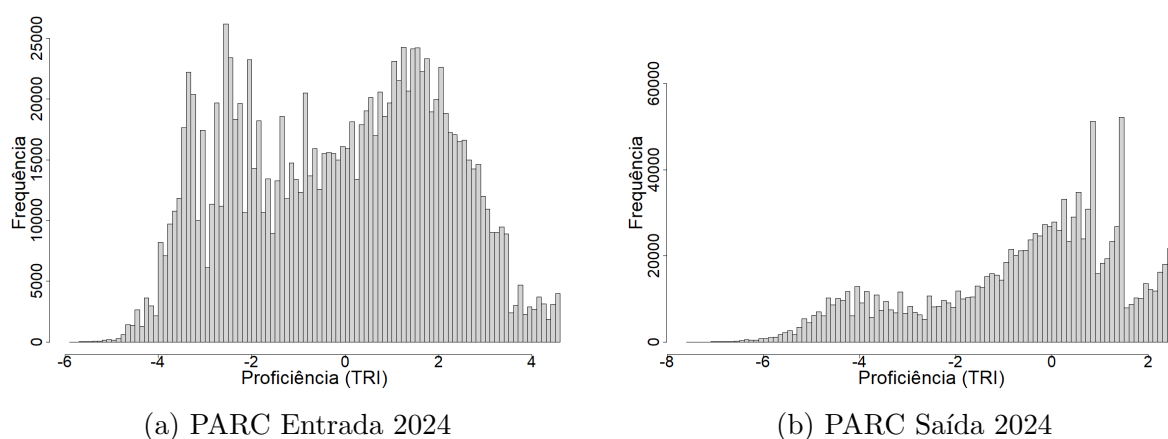


Figura 5.6: Distribuição da proficiência estimada pela TRI. Os gráficos utilizam escalas independentes nos eixos horizontal e vertical para facilitar a visualização das distribuições.

Fonte: Elaborado pelo autor.

Observa-se um deslocamento da distribuição de proficiência em direção a valores mais elevados na avaliação de Saída quando comparada à avaliação de Entrada. Enquanto a distribuição inicial apresenta maior concentração de estudantes em níveis mais baixos de habilidade, a distribuição observada ao final do ano letivo concentra maior proporção de estudantes em níveis intermediários e superiores da escala. Esse comportamento sugere desenvolvimento das habilidades de leitura ao longo do período analisado, refletindo o avanço esperado dos estudantes durante o ano escolar.

Comportamento semelhante pode ser observado nas distribuições de QPC, apre-

sentadas na Figura 5.7. Na avaliação de Entrada, verifica-se elevada concentração de estudantes com baixos valores de QPC, indicando que uma parcela significativa dos participantes ainda apresentava dificuldades na leitura das palavras avaliadas. Já na avaliação de Saída, observa-se maior concentração de estudantes nos níveis superiores da escala, evidenciando aumento do desempenho observado e corroborando os resultados obtidos por meio das estimativas de proficiência.

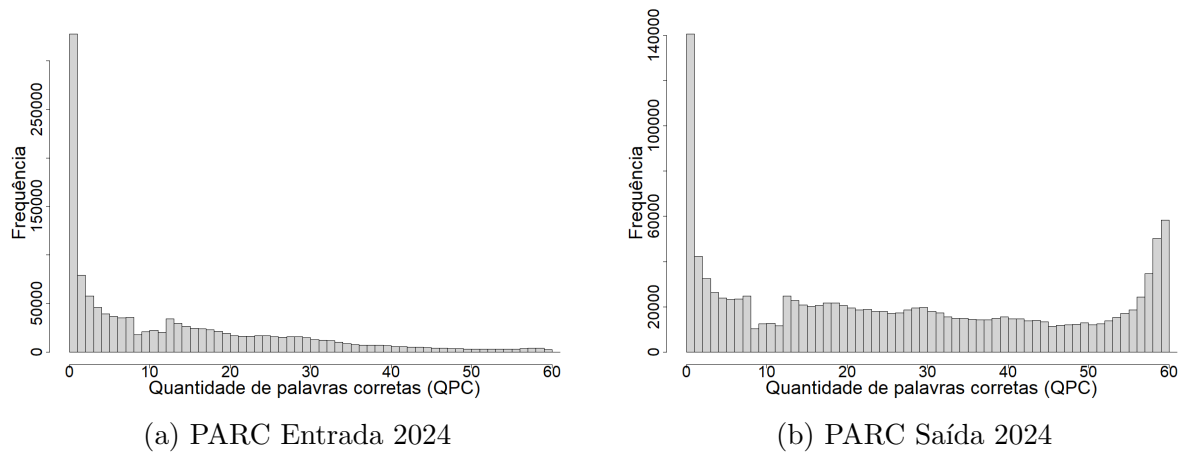


Figura 5.7: Distribuição de quantidade de palavras corretas. Os gráficos utilizam escalas independentes no eixo vertical para facilitar a visualização das distribuições.

Fonte: Elaborado pelo autor.

Além da análise isolada da distribuição das proficiências, buscou-se verificar em que medida os níveis de dificuldade dos itens utilizados nas avaliações correspondem às diferentes regiões da distribuição de habilidade dos estudantes. Essa comparação é possível porque, no Modelo de Rasch, os parâmetros de proficiência dos indivíduos (θ) e de dificuldade dos itens (b) são estimados em uma mesma escala métrica. Dessa forma, a distribuição dos estudantes e dos itens ao longo dessa escala permite avaliar quais regiões do continuum de proficiência são mais ou menos representadas pelo instrumento.

Para facilitar a comparação, a escala foi dividida em cinco faixas definidas a partir da média e do desvio-padrão das proficiências estimadas em cada aplicação: valores inferiores a $-1,5$ desvio-padrão, entre $-1,5$ e $-0,5$, entre $-0,5$ e $0,5$, entre $0,5$ e $1,5$ e superiores a $1,5$ desvio-padrão. Os limites numéricos obtidos a partir da distribuição das proficiências foram também aplicados aos parâmetros de dificuldade dos itens, permitindo calcular, em cada faixa, o percentual de estudantes e o percentual de itens. A Figura 5.8

apresenta essa comparação para as avaliações PARC Entrada 2024 e PARC Saída 2024.

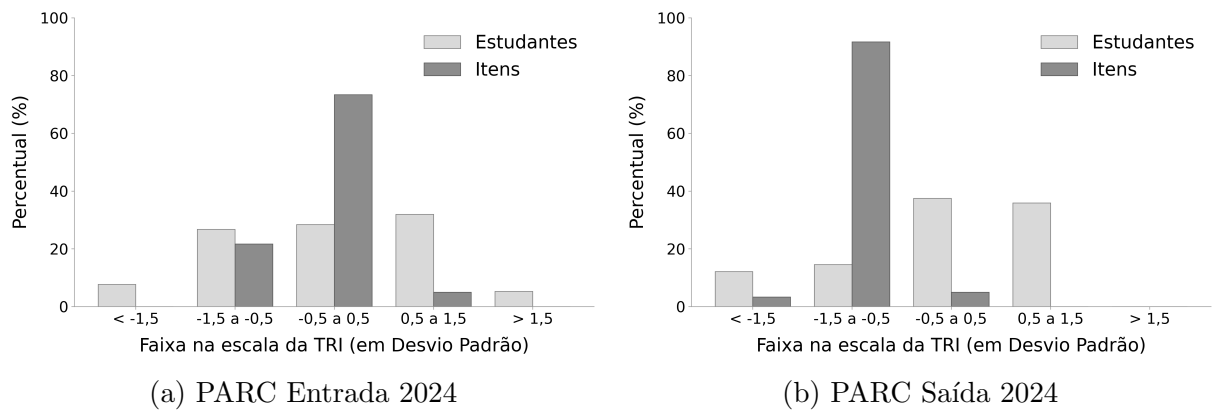


Figura 5.8: Distribuição percentual dos estudantes e dos itens por faixas da escala da TRI nas avaliações PARC Entrada 2024 e PARC Saída 2024. As faixas foram definidas a partir da média e do desvio-padrão das proficiências estimadas em cada aplicação.

Fonte: Elaborado pelo autor.

A comparação entre a distribuição das proficiências dos estudantes e das dificuldades dos itens evidencia diferenças na cobertura da escala entre as duas aplicações. Na PARC Entrada 2024, as dificuldades dos itens concentram-se predominantemente na faixa central, entre $-0,5$ e $0,5$ desvio-padrão, enquanto as proficiências dos estudantes apresentam maior dispersão entre as diferentes regiões da escala. Na PARC Saída 2024, observa-se um deslocamento das proficiências para faixas mais elevadas, enquanto a maior parte dos itens se concentra entre $-1,5$ e $-0,5$ desvio-padrão. Além disso, não são observados itens nas faixas superiores da escala, apesar da presença de uma parcela expressiva de estudantes nesses níveis de proficiência.

Esse resultado indica que, especialmente na avaliação de saída, a dificuldade dos itens se concentra abaixo da proficiência de parte considerável dos participantes, sugerindo menor cobertura das regiões superiores da escala. A inclusão de palavras com maior dificuldade poderia, portanto, ampliar a cobertura do instrumento e favorecer a diferenciação entre estudantes com níveis mais elevados de proficiência.

A Figura 5.9 apresenta a relação entre a quantidade de palavras corretas e a proficiência estimada pela TRI. Em ambas as aplicações observa-se uma associação positiva entre as variáveis, indicando que estudantes com maior número de acertos tendem a apresentar maiores níveis de habilidade. Os coeficientes de correlação de Pearson foram iguais

a 0,8188 para a PARC Entrada 2024 e 0,8246 para a PARC Saída 2024, caracterizando associação forte entre os indicadores nos dois momentos avaliados.

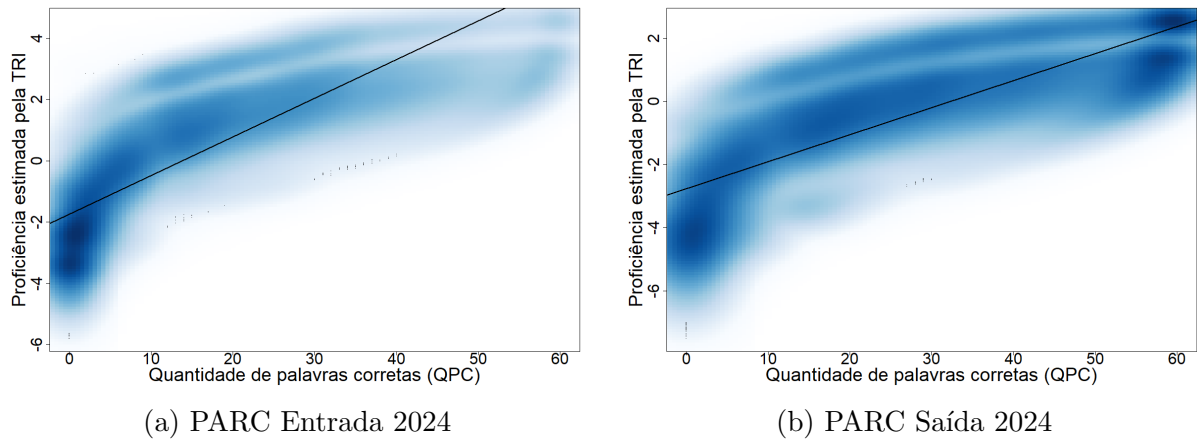


Figura 5.9: Relação entre QPC e proficiência TRI. Os gráficos utilizam escalas independentes no eixo vertical para facilitar a visualização das distribuições.

Fonte: Elaborado pelo autor.

Embora a relação entre as variáveis seja claramente crescente, observa-se dispersão dos valores de proficiência para determinados níveis de QPC. Esse resultado era esperado, uma vez que a proficiência estimada pela TRI não depende exclusivamente da quantidade total de acertos, mas também da dificuldade dos itens respondidos e do padrão de respostas apresentado por cada estudante. Dessa forma, estudantes com a mesma quantidade de palavras corretas podem apresentar níveis de proficiência distintos, dependendo das palavras acertadas e erradas ao longo da avaliação.

A Figura 5.10 apresenta a proficiência média estimada para cada valor possível de QPC. Em ambas as avaliações observa-se crescimento monotônico da proficiência média à medida que a quantidade de palavras corretas aumenta, reforçando a consistência entre a medida psicométrica produzida pela TRI e a métrica tradicional baseada na contagem de acertos. Observa-se ainda que esse crescimento não ocorre de forma estritamente linear, especialmente nos extremos da escala, evidenciando que incrementos semelhantes na quantidade de acertos podem produzir ganhos distintos de proficiência.

Os modelos de regressão linear simples ajustados para as duas avaliações apresentaram coeficientes de determinação semelhantes, indicando que aproximadamente dois terços da variabilidade observada nas proficiências podem ser explicados pela quantidade

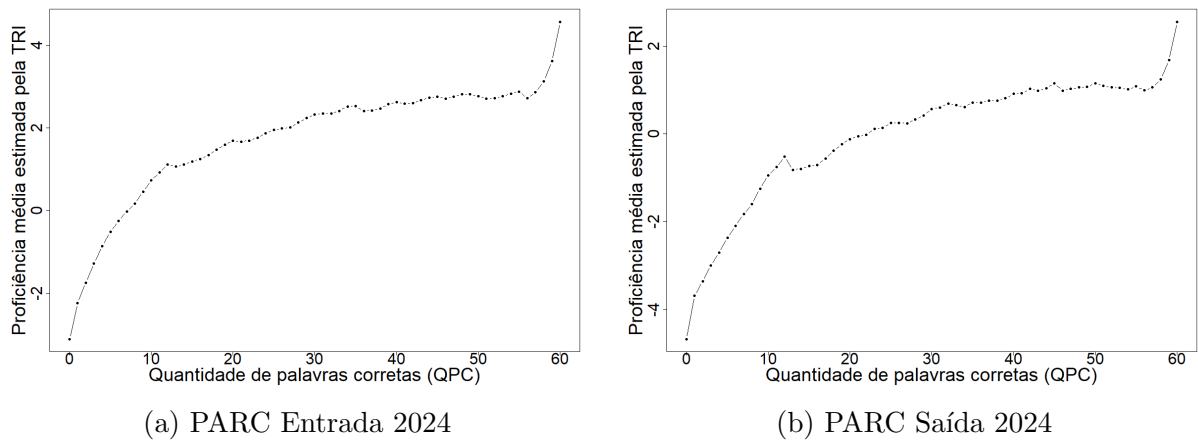


Figura 5.10: Proficiência média estimada pela TRI em função da quantidade de palavras corretas (QPC). Os gráficos utilizam escalas independentes no eixo vertical para facilitar a visualização das distribuições.

Fonte: Elaborado pelo autor.

de palavras corretas. Em ambos os casos, os coeficientes associados à QPC foram estatisticamente significativos ($p < 0,001$), corroborando a existência de uma relação forte entre as duas medidas.

Tabela 5.3: Resultados da correlação e da regressão linear simples entre QPC e proficiência estimada

Avaliação	Correlação	R^2
PARC Entrada 2024	0,8188	0,6705
PARC Saída 2024	0,8246	0,6800

Fonte: Elaborado pelo autor.

Em conjunto, os resultados indicam que as proficiências estimadas pela Teoria de Resposta ao Item apresentam elevada concordância com a quantidade de palavras corretas tanto na avaliação de Entrada quanto na avaliação de Saída. Além disso, a comparação entre os dois momentos evidencia um deslocamento consistente das distribuições de desempenho para níveis mais elevados, sugerindo desenvolvimento das habilidades de leitura dos estudantes ao longo do ano letivo. Os resultados também demonstram que a proficiência estimada pela TRI fornece uma representação coerente do desempenho observado, incorporando simultaneamente a quantidade de acertos e a dificuldade dos itens respondidos.

5.4 Análise da Dificuldade das Palavras

Com o objetivo de investigar quais características das palavras estão relacionadas com a dificuldade estimada pela Teoria de Resposta ao Item (TRI), foram ajustados modelos de regressão linear múltipla estimados pelo método dos Mínimos Quadrados Ordinários (*Ordinary Least Squares* – OLS) para as avaliações PARC Entrada 2024 e PARC Saída 2024. Como variável dependente foi utilizada a dificuldade estimada dos itens pela TRI, enquanto as variáveis independentes corresponderam às características linguísticas das palavras e às características relacionadas à sua distribuição nos cadernos de aplicação.

Dentre as variáveis consideradas, estão características consideradas no padrão técnico de montagem do teste, sendo elas: número de sílabas, canonicidade, ocorrência das ambiguidades ortográficas envolvendo os fonemas /s/-/z/ e /g/-/j/, tonicidade. As outras variáveis utilizadas foram: posição média da palavra nos cinco cadernos, desvio padrão da posição da palavra entre os cadernos, frequência lexical e número de letras.

A Tabela 5.4 apresenta as variáveis independentes utilizadas nos modelos de regressão linear múltipla, bem como sua natureza e domínio.

Tabela 5.4: Variáveis independentes utilizadas no modelo de regressão linear múltipla.

Característica	Natureza	Domínio
Número de sílabas	Ordinal	1, 2, 3, 4
Canonicidade	Binária	0, 1
Ambiguidades ortográficas - fonemas /s/ e /z/	Binária	0, 1
Ambiguidades ortográficas - fonemas /g/ e /j/	Binária	0, 1
Tonicidade	Ordinal	1, 2, 3
Posição média da palavra nos cadernos	Quantitativa contínua	$\mathbb{R}$
Desvio-padrão da posição da palavra nos cadernos	Quantitativa contínua	$\mathbb{R}_{\geq 0}$
Frequência lexical	Quantitativa discreta	$\mathbb{N}$
Número de letras	Quantitativa discreta	$\mathbb{N}$

Fonte: Elaborado pelo autor.

A Figura 5.11 apresenta a comparação entre a dificuldade observada pela TRI e a dificuldade estimada pelos modelos de regressão linear múltipla para as avaliações PARC Entrada 2024 e PARC Saída 2024, respectivamente.

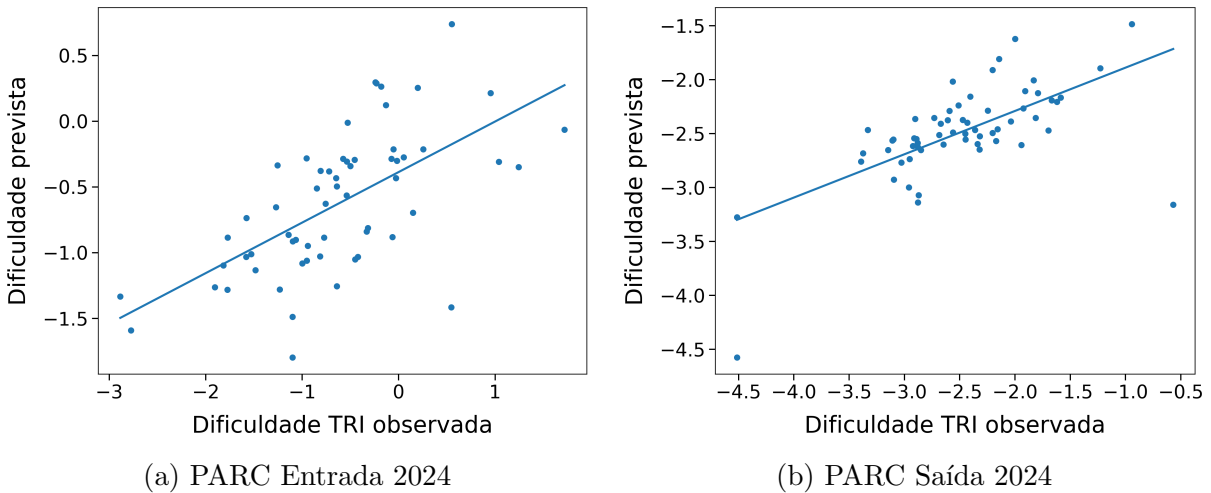


Figura 5.11: Dificuldade prevista X calculada pela TRI.

Fonte: Elaborado pelo autor.

Os resultados obtidos indicam que o modelo foi capaz de explicar aproximadamente 38,4% da variabilidade da dificuldade dos itens na PARC Entrada ($R^2 = 0,384$) e 40,3% na PARC Saída ($R^2 = 0,403$). Após o ajuste pelo número de variáveis do modelo, os coeficientes de determinação ajustados foram de 0,274 e 0,295, respectivamente. Embora esses valores indiquem que parte da variabilidade permanece não explicada, eles demonstram que uma parcela considerável da dificuldade dos itens pode ser atribuída às características estruturais das palavras.

Além do coeficiente de determinação, foram calculadas medidas de associação entre os valores observados pela TRI e os valores previstos pela regressão linear múltipla. Na PARC Entrada foram obtidos coeficientes de Pearson de 0,620, Spearman de 0,665 e Kendall τ de 0,487. Na PARC Saída esses valores aumentaram para 0,635, 0,679 e 0,501, respectivamente. Esses resultados indicam associação moderada entre os valores observados e previstos, evidenciando que o modelo consegue reproduzir satisfatoriamente a ordenação relativa da dificuldade dos itens.

A Tabela 5.5 apresenta um resumo das principais métricas obtidas para ambos os modelos.

Tabela 5.5: Comparação entre os modelos de regressão linear múltipla para a PARC Entrada e PARC Saída

Métrica	PARC Entrada	PARC Saída
R^2	0,384	0,403
R^2 Ajustado	0,274	0,295
Correlação de Pearson	0,620	0,635
Correlação de Spearman	0,665	0,679
Kendall τ	0,487	0,501

Observa-se que o modelo apresentou desempenho bastante semelhante nas duas aplicações da avaliação. A pequena melhora observada na PARC Saída pode estar relacionada ao desenvolvimento das habilidades de leitura dos estudantes ao longo do ano letivo. À medida que os alunos adquirem maior fluência leitora, espera-se que características intrínsecas das palavras, como frequência de ocorrência e extensão, passem a exercer influência mais consistente sobre a dificuldade observada, reduzindo parte da variabilidade

decorrente das diferenças individuais entre os estudantes.

A análise dos coeficientes da regressão linear múltipla revelou diferenças importantes entre as duas aplicações. A Tabela 5.6 resume os coeficientes estimados e sua significância estatística.

Tabela 5.6: Comparação dos coeficientes estimados para os modelos de regressão linear múltipla

Variável	PARC Entrada	PARC Saída
Tipo silábico	Não significativo	Tendência ($p = 0,091$)
Canonicidade	Não significativo	Não significativo
Variações /s/ e /z/	Não significativo	Não significativo
Concorrência /g/ e /j/	Não significativo	Não significativo
Tonicidade	Não significativo	Não significativo
Posição média	Não significativo	Não significativo
Desvio padrão da posição	Não significativo	Não significativo
Frequência lexical	Significativo ($p = 0,032$)	Significativo ($p = 0,005$)
Número de letras	Não significativo	Significativo ($p = 0,032$)

Dentre todas as variáveis analisadas, a frequência lexical foi a única que apresentou significância estatística em ambas as aplicações. O coeficiente negativo obtido indica que palavras mais frequentes tendem a apresentar menor dificuldade estimada pela TRI, resultado consistente com a literatura sobre reconhecimento de palavras e processamento lexical, segundo a qual palavras mais frequentes são acessadas mais rapidamente pelos leitores.

Na PARC Saída, o número de letras também apresentou associação significativa com a dificuldade dos itens. Esse resultado sugere que, após um período de desenvolvimento das habilidades de leitura, o comprimento da palavra passa a exercer influência adicional sobre sua dificuldade. Em contrapartida, essa variável não apresentou significância na PARC Entrada, indicando que, nos estágios iniciais da aprendizagem, outros fatores relacionados ao próprio processo de alfabetização podem exercer maior influência sobre o desempenho dos estudantes.

Por outro lado, as variáveis relacionadas às ambiguidades ortográficas (/s/-/z/

e /g/-/j/), à canonicidade, à tonicidade e à posição das palavras nos diferentes cadernos não apresentaram efeito estatisticamente significativo quando analisadas em conjunto com as demais características. Esse resultado não implica necessariamente que essas características não influenciem a dificuldade das palavras, mas indica que sua contribuição adicional torna-se reduzida após o controle das variáveis de frequência lexical e comprimento da palavra.

Durante a construção dos modelos também foi avaliada a presença de multicolinearidade entre as variáveis independentes por meio do Fator de Inflação da Variância (Variance Inflation Factor – VIF). Observou-se elevada colinearidade entre o número de sílabas e o número de letras, indicando que ambas representam, em grande medida, uma mesma dimensão relacionada ao comprimento da palavra. As demais variáveis apresentaram valores de VIF inferiores a três, indicando baixa colinearidade e boa estabilidade dos coeficientes estimados.

De forma geral, os resultados obtidos indicam que características objetivas das palavras são capazes de explicar uma parcela importante da dificuldade estimada pela TRI. Entretanto, aproximadamente 60% da variabilidade permanece não explicada pelos modelos, sugerindo que outros fatores, como idade de aquisição, concretude, familiaridade, imageabilidade e aspectos individuais dos estudantes, também podem desempenhar papel relevante na determinação da dificuldade dos itens.

5.4.1 Estimativa da Proficiência dos Estudantes a partir das Dificuldades Preditas

Como experimento complementar, foi investigado se as dificuldades dos itens estimadas pelo modelo de regressão linear múltipla seriam capazes de produzir estimativas de proficiência semelhantes às obtidas utilizando as dificuldades calibradas diretamente pela Teoria de Resposta ao Item (TRI).

Para isso, as dificuldades previstas pelo modelo de regressão linear múltipla foram utilizadas como parâmetros fixos dos itens em um modelo de Rasch (1PL), mantendo-se inalteradas as matrizes de respostas dos estudantes. Esse procedimento foi realizado separadamente para as avaliações PARC Entrada 2024 e PARC Saída 2024, produzindo

uma nova estimativa da proficiência (θ) de cada estudante. Posteriormente, essas novas estimativas foram comparadas às proficiências originalmente obtidas pela calibração da TRI.

A Figura 5.12 apresenta a relação entre as proficiências estimadas pela TRI e aquelas obtidas utilizando as dificuldades previstas pelo modelo de regressão linear múltipla.

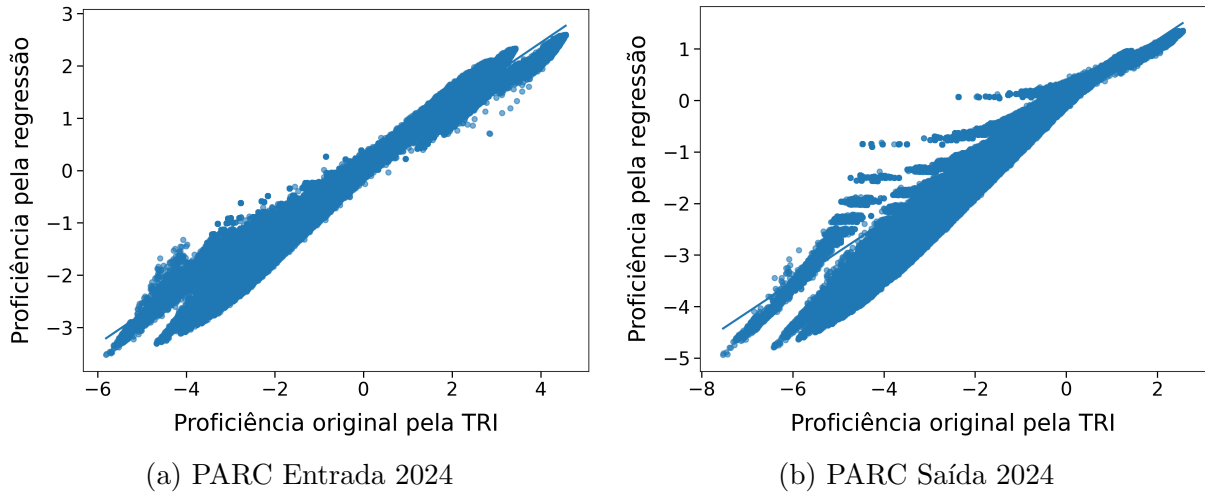


Figura 5.12: Comparação entre a proficiência estimada pela TRI e a proficiência estimada utilizando as dificuldades previstas pelo modelo de regressão linear múltipla.

Fonte: Elaborado pelo autor.

Observa-se que, tanto na avaliação de Entrada quanto na de Saída, existe forte associação linear entre as duas estimativas de proficiência. Em ambas as aplicações, a maior parte dos estudantes concentra-se próxima da reta de regressão, indicando elevada concordância entre as proficiências obtidas utilizando os parâmetros calibrados pela TRI e aquelas estimadas a partir das dificuldades previstas pelo modelo de regressão linear múltipla.

Para quantificar essa relação foram calculados os coeficientes de correlação de Pearson e Spearman, além da raiz do erro quadrático médio (RMSE) e do erro médio absoluto (MAE). Os resultados encontram-se resumidos na Tabela 5.7.

Tabela 5.7: Comparação entre as proficiências estimadas pela TRI e pela regressão linear múltipla

Métrica	PARC Entrada	PARC Saída
Correlação de Pearson	0,983	0,967
Correlação de Spearman	0,985	0,990
RMSE	0,968	0,943
MAE	0,754	0,671

Os resultados evidenciam elevada concordância entre as duas formas de estimação da proficiência. Em ambas as aplicações foram obtidas correlações de Pearson superiores a 0,96, indicando que as proficiências calculadas utilizando as dificuldades previstas pela regressão linear múltipla apresentam comportamento muito semelhante às proficiências originalmente estimadas pela TRI. Os elevados coeficientes de Spearman reforçam essa conclusão ao demonstrar que a ordenação relativa dos estudantes foi praticamente preservada.

Observa-se ainda que a avaliação PARC Saída apresentou desempenho ligeiramente superior em relação à PARC Entrada, apresentando menor RMSE e menor MAE, além de maior correlação de Spearman. Esse comportamento sugere que, à medida que os estudantes desenvolvem suas habilidades de leitura ao longo do ano letivo, a dificuldade dos itens passa a ser explicada de forma mais consistente pelas características linguísticas das palavras consideradas no modelo de regressão linear múltipla.

Esse resultado é coerente com as análises apresentadas na subseção anterior, nas quais a frequência lexical e o comprimento das palavras demonstraram maior capacidade de explicar a dificuldade estimada pela TRI. Assim, espera-se que essas características também contribuam para uma estimação mais estável da proficiência quando utilizadas como parâmetros dos itens.

É importante destacar que esse experimento não tem como objetivo substituir a calibração tradicional realizada pela TRI. As dificuldades previstas pela regressão linear múltipla foram obtidas a partir de modelos previamente ajustados utilizando dificuldades calibradas pela própria TRI, de modo que ambas as abordagens permanecem conceitualmente relacionadas. Dessa forma, os resultados devem ser interpretados como uma va-

validação da capacidade do modelo de regressão linear múltipla em aproximar os parâmetros dos itens, e não como uma alternativa definitiva ao processo de calibração psicométrica.

Ainda assim, os resultados obtidos demonstram o potencial de utilização do modelo proposto como ferramenta auxiliar na construção de instrumentos de avaliação. Em situações nas quais novos conjuntos de palavras ainda não disponham de respostas suficientes para uma calibração completa pela TRI, as dificuldades estimadas a partir das características linguísticas podem fornecer valores iniciais bastante consistentes para os parâmetros dos itens. A elevada concordância observada entre as proficiências estimadas pelas duas abordagens indica que essa estratégia constitui uma alternativa promissora, cuja aplicação em novos conjuntos de itens e em diferentes populações poderá ser investigada em trabalhos futuros.

5.5 Discussão

Os resultados obtidos neste trabalho evidenciam que a Teoria de Resposta ao Item mostrou-se adequada para a modelagem da dificuldade de leitura de palavras isoladas em língua portuguesa. A verificação da unidimensionalidade indicou a existência de um fator dominante nas avaliações analisadas, atendendo ao principal pressuposto do Modelo de Rasch e permitindo que as estimativas de proficiência dos estudantes e de dificuldade dos itens fossem obtidas em uma mesma escala latente.

A análise da distribuição das dificuldades entre os diferentes cadernos também fornece evidências favoráveis à construção do instrumento. Apesar de os cinco cadernos apresentarem ordens distintas de apresentação das palavras, observou-se que a dificuldade acumulada permaneceu bastante semelhante entre eles, indicando que a distribuição dos itens foi suficientemente equilibrada. Esse resultado é particularmente relevante, pois reduz a possibilidade de que diferenças de desempenho entre estudantes sejam decorrentes da versão do caderno recebida, aumentando a comparabilidade das estimativas de proficiência.

Outro resultado importante foi a elevada associação observada entre a proficiência estimada pela TRI e a quantidade de palavras corretas (QPC). Embora a QPC constitua uma medida simples e amplamente utilizada na prática educacional, a TRI oferece uma

representação mais completa do desempenho do estudante ao considerar simultaneamente sua habilidade e a dificuldade dos itens respondidos. Dessa forma, a forte correlação observada sugere que ambas as medidas refletem o mesmo construto de leitura, ao mesmo tempo em que evidencia a vantagem da TRI por produzir estimativas independentes do conjunto específico de itens utilizado.

Em relação às características das palavras, os modelos de regressão indicaram que variáveis linguísticas e contextuais apresentam associação com a dificuldade estimada pela TRI. Entre os preditores identificados, destacam-se características relacionadas à estrutura das palavras, ao seu tamanho e à frequência lexical delas. Esses resultados reforçam a hipótese de que a dificuldade de leitura não depende exclusivamente da habilidade do estudante, mas também de propriedades intrínsecas dos itens avaliados.

Esses achados são consistentes com a literatura apresentada no Capítulo 3. Athayde et al. (2019) destacam a importância da seleção criteriosa de características linguísticas na construção de instrumentos de leitura, enquanto Compton et al. (2025) demonstram que propriedades lexicais contribuem significativamente para explicar a dificuldade das palavras. Embora o presente trabalho tenha utilizado um conjunto diferente de características, os resultados convergem ao evidenciar que atributos linguísticos podem auxiliar na compreensão da dificuldade dos itens e na construção de avaliações mais equilibradas.

Entretanto, observa-se que o poder explicativo dos modelos de regressão permaneceu moderado. Esse resultado sugere que a dificuldade de leitura é influenciada por múltiplos fatores, muitos dos quais não foram contemplados neste estudo. Características como idade de aquisição da palavra, transparência ortográfica, familiaridade lexical e variáveis semânticas, frequentemente apontadas na literatura internacional, podem contribuir para explicar parte da variabilidade não capturada pelos modelos desenvolvidos.

A estimativa das proficiências utilizando dificuldades previstas pela regressão linear múltipla também apresentou resultados relevantes. Apesar de as dificuldades estimadas pelo modelo não substituírem completamente aquelas obtidas diretamente pela TRI, verificou-se elevada concordância entre as proficiências calculadas pelos dois métodos. Esse resultado demonstra que modelos baseados em características linguísticas podem produzir aproximações consistentes da dificuldade dos itens, representando uma alterna-

tiva promissora para auxiliar a elaboração de novos instrumentos antes mesmo de sua aplicação em larga escala.

Do ponto de vista prático, os resultados desta pesquisa podem contribuir para diferentes etapas da construção de avaliações de fluência leitora. As estimativas de dificuldade obtidas pela TRI permitem selecionar palavras de diferentes níveis de complexidade, balancear a distribuição dos itens entre diferentes versões de um mesmo instrumento e produzir escalas de proficiência mais comparáveis ao longo do tempo. Além disso, os modelos de regressão desenvolvidos podem auxiliar na seleção inicial de palavras para novos testes, reduzindo a dependência exclusiva da avaliação subjetiva por especialistas.

Por fim, os resultados também evidenciam algumas limitações da pesquisa. O estudo foi desenvolvido utilizando exclusivamente o Modelo de Rasch, que considera apenas o parâmetro de dificuldade dos itens. Além disso, foram analisadas apenas avaliações de leitura de palavras isoladas, não sendo considerados outros aspectos da fluência leitora, como velocidade, prosódia e compreensão textual. Essas limitações indicam possibilidades para pesquisas futuras, incluindo a utilização de modelos mais complexos da Teoria de Resposta ao Item e a incorporação de novas características linguísticas capazes de ampliar o poder preditivo dos modelos desenvolvidos.

6 Conclusões

Este trabalho teve como objetivo utilizar a Teoria de Resposta ao Item, por meio do Modelo de Rasch, para estimar a dificuldade de leitura de palavras isoladas em língua portuguesa a partir de dados das avaliações PARC Entrada 2024 e PARC Saída 2024. Além da estimação dos parâmetros de dificuldade dos itens e da proficiência dos estudantes, buscou-se investigar a distribuição dessas dificuldades nos diferentes cadernos de avaliação, analisar sua relação com características linguísticas das palavras e verificar a possibilidade de prever dificuldades por meio de modelos de regressão.

Os resultados obtidos permitiram atingir os objetivos propostos. A aplicação do Modelo de Rasch possibilitou estimar, em uma mesma escala latente, a dificuldade das palavras e a proficiência dos estudantes, produzindo medidas que se mostraram consistentes para a análise do instrumento. A verificação das premissas do modelo indicou que os dados apresentavam condições adequadas para sua utilização, fornecendo suporte às análises realizadas ao longo do trabalho.

Também foi possível demonstrar que os diferentes cadernos da avaliação apresentam distribuição equilibrada de dificuldade, apesar das diferenças na ordem de apresentação das palavras. Esse resultado contribui para reforçar a equivalência entre as versões do instrumento e reduz a possibilidade de que diferenças de desempenho sejam atribuídas ao caderno recebido pelo estudante.

A investigação das características linguísticas evidenciou que parte da dificuldade das palavras pode ser explicada por atributos estruturais e contextuais, permitindo a construção de modelos capazes de prever, ainda que parcialmente, a dificuldade dos itens. Além disso, verificou-se que as proficiências estimadas utilizando essas dificuldades preditas apresentaram elevada concordância com aquelas obtidas a partir das dificuldades calibradas diretamente pela TRI, indicando que modelos desse tipo podem auxiliar no planejamento e na construção de novos instrumentos de avaliação.

Como principal contribuição, esta pesquisa apresenta uma abordagem integrada para a análise da dificuldade de palavras em avaliações de fluência leitora, reunindo pro-

cedimentos de calibração psicométrica, análise da distribuição dos itens, investigação de características linguísticas e modelagem estatística da dificuldade.

Entretanto, algumas limitações devem ser consideradas. O estudo foi desenvolvido utilizando exclusivamente o Modelo de Rasch, que considera apenas o parâmetro de dificuldade dos itens. Além disso, as análises contemplaram um conjunto específico de características linguísticas, não incluindo variáveis potencialmente relevantes apontadas pela literatura, como idade de aquisição, transparência ortográfica e medidas de familiaridade lexical. Da mesma forma, a pesquisa concentrou-se em avaliações de leitura de palavras isoladas, não contemplando outros componentes da fluência leitora, como velocidade, prosódia e compreensão.

Como perspectivas para trabalhos futuros, sugere-se ampliar o conjunto de características linguísticas utilizadas na modelagem da dificuldade das palavras, investigar a aplicação de modelos mais complexos da Teoria de Resposta ao Item, como o modelo de dois parâmetros, bem como modelos multidimensionais, e avaliar o comportamento dessas abordagens em diferentes populações e níveis de escolaridade. Também se mostra promissora a utilização dos modelos desenvolvidos para apoiar a seleção automática de palavras e a construção de novos instrumentos de avaliação da leitura.

Espera-se, por fim, que os resultados apresentados contribuam para o desenvolvimento de avaliações de fluência leitora mais fundamentadas em evidências empíricas, auxiliando pesquisadores e profissionais da educação na elaboração de instrumentos capazes de mensurar, de forma mais precisa, o desenvolvimento da leitura em língua portuguesa.

Bibliografia

- ANDRADE, D. F. d.; TAVARES, H. R.; VALLE, R. d. C. *Teoria da Resposta ao Item: conceitos e aplicações*. São Paulo: Associação Brasileira de Estatística, 2000.
- ANDRADE, J. M. d.; LAROS, J. A.; GOUVEIA, V. V. O uso da teoria de resposta ao item em avaliações educacionais: diretrizes para pesquisadores. *Avaliação Psicológica*, v. 9, n. 3, p. 421–435, 2010.
- ATHAYDE, M. D. L. et al. Desenvolvimento do subteste de leitura do teste de desempenho escolar ii. *Psico-USF*, v. 24, n. 2, p. 245–257, 2019.
- BECKER, F. d. R. Avaliação educacional em larga escala: a experiência brasileira. *Revista Iberoamericana de Educación*, v. 53, n. 1, p. 1–13, 2010.
- BLOOM, B. S.; HASTINGS, J. T.; MADDAUS, G. F. *Handbook on Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill, 1971.
- BONAMINO, A.; SOUSA, S. Z. Três gerações de avaliação da educação básica no brasil: interfaces com o currículo da/na escola. *Educação e Pesquisa*, v. 38, n. 2, p. 373–388, 2012.
- Brasil. Ministério da Educação. *Base Nacional Comum Curricular*. Brasília: MEC, 2018.
- COMPTON, D. L. et al. The developmental english lexicon project (d-elp): Continuous word reading difficulty estimates of the 9,961 most frequently printed english words for us children. *Behavior Research Methods*, 2025. Under review.
- DANTAS, L. A. Avaliação e verificação escolar: uma análise histórica das legislações educacionais do brasil. *História & Ensino*, v. 29, p. 157–182, 2023.
- DUBECK, M. M.; GOVE, A. The early grade reading assessment (egra): Its theoretical foundation, purpose, and limitations. *International Journal of Educational Development*, v. 40, p. 315–322, 2015.
- EHRI, L. C. Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, v. 9, n. 2, p. 167–188, 2005.
- FERREIRA, F. F. G. *Escala de Proficiência para o ENEM utilizando a Teoria da Resposta ao Item*. Dissertação (Dissertação (Mestrado em Matemática e Estatística)) — Universidade Federal do Pará, Belém, 2009.
- HACHÉN, R. Conocimiento lingüístico y reflexión metalingüística: El rol de las conceptualizaciones en torno a la estructura de la sílaba en el proceso de alfabetización. *Lectura y Vida*, v. 23, n. 2, p. 6–19, 2002.
- HAMBLETON, R. K.; SWAMINATHAN, H.; ROGERS, H. J. *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publications, 1991.
- KNIJNIK, L. F. et al. Avaliação dos subtestes de leitura e escrita do teste de desempenho escolar através da teoria de resposta ao item. *Psicologia: Reflexão e Crítica*, v. 27, n. 3, p. 481–490, 2014.

LÚCIO, P. S. et al. Word decoding task: item analysis by irt and within-group norms. *Psicologia: Teoria e Pesquisa*, v. 34, p. e3437, 2018.

PASQUALI, L. *Psicometria: teoria dos testes na psicologia e na educação*. 5. ed. Petrópolis: Vozes, 2003. 5. ed. publicada em 2017.

PASQUALI, L.; PRIMI, R. Fundamentos da teoria da resposta ao item: Tri. *Avaliação Psicológica*, v. 2, n. 2, p. 99–110, 2003.

SEYMOUR, P. H. K.; ARO, M.; ERSKINE, J. M. Foundation literacy acquisition in european orthographies. *British Journal of Psychology*, v. 94, n. 2, p. 143–174, 2003.

SILVA, J. T. F.; SILVA, W.; MEGHERBI, H. Construção e validação de instrumentos para avaliação da fluência em leitura. In: *Anais da XII Reunião da Associação Brasileira de Avaliação Educacional (ABAVE)*. Even3, 2023. ISBN 978-65-272-0302-5. Disponível em: <https://www.even3.com.br/anais/xiiabave/661250-construcao-e-validacao-de-instrumentos-para-avaliacao-da-fluencia-em-leitura/>.

TEDESCO, J. C. Evaluar las evaluaciones. In: *Evaluar las Evaluaciones: una mirada política acerca de las evaluaciones de la calidad educativa*. Buenos Aires, Argentina: IIPE–UNESCO, 2003.