

Fidedignidade, validade interna e normatização dos testes de reconhecimento de palavras e de pseudopalavras

Fiabilidad, validez interna y estandarización de los tests de reconocimiento de palabras y pseudopalabras
Fiabilité, validité interne et standardisation des tests de reconnaissance de mots et de pseudo-mots
Reliability, Internal Validity and Standardization of the Words and the Pseudowords Recognition Tests

Douglas de Araújo Vilhena¹, Ângela Maria Vieira Pinheiro¹

1. Programa de Pós-Graduação em Psicologia: Cognição e Comportamento, Faculdade de Filosofia e Ciências Humanas, Universidade Federal de Minas Gerais, Brasil.

Resumo

A presente pesquisa objetivou provar as fontes de evidências psicométricas para o 'Teste de Reconhecimento de Palavras' e o 'Teste de Reconhecimento de Pseudopalavras' no que concerne a: (a) fidedignidade; (b) validade da estrutura interna; e (c) normatização. Participaram 598 alunos, 52% do sexo masculino, do 2º ao 5º ano do Ensino Fundamental, de 7 a 11 anos, estratificados em 102 salas de aula de oito escolas estaduais. Os escores de ambos os instrumentos são dados em Acurácia (porcentagens de palavras lidas corretamente) e Taxa de Acurácia (número de palavras lidas corretamente por minuto). Como evidência de fidedignidade, houve consistência interna alta e equivalência entre as duas metades e as formas paralelas. Validade da Estrutura Interna foi verificada com o efeito de escolaridade, com a variável Taxa de Acurácia apresentado melhores índices psicométricos do que a Acurácia, com curva de distribuição normal e sem sobreposição dos Intervalos de Confiança entre os anos escolares. Concluiu-se que o Teste de Reconhecimento de Palavras e o Teste de Reconhecimento de Pseudopalavras demonstraram evidências psicométricas satisfatórias de fidedignidade e validade interna, o que sustenta o estabelecimento de normas padronizadas para o 2º ao 5º ano escolar.

Palavras-chave: deficiências da aprendizagem, dislexia, estudo de validação, leitura, psicomетria.

Resumen

La presente investigación tuvo como objetivo proporcionar las fuentes de evidencias psicométricas para el "Test de reconocimiento de palabras" y el "Test de reconocimiento de pseudopalabras" en lo relacionado a: (a) fiabilidad; (b) validez de la estructura interna; y (c) normatización. Los participantes fueron 598 alumnos, 52% del sexo masculino, de 2º a 5º de Primaria, de 7 a 11 años, estratificados en 102 aulas de ocho colegios públicos. Las puntuaciones de ambos instrumentos son dados en Exactitud (porcentajes de palabras leídas correctamente) y Tasa de Exactitud (número de palabras leídas correctamente por minuto). Como evidencia de la fiabilidad, hubo una alta consistencia interna y equivalencia entre las dos mitades y las formas paralelas. Se verificó la Validez de la Estructura Interna con el efecto de la escolaridad, mostrando la variable Tasa de Exactitud mejores índices psicométricos que la Exactitud, con curva de distribución normal y sin superposición de Intervalos de Confianza entre los años escolares. Se concluyó que el Test de Reconocimiento de Palabras y el Test de Reconocimiento de Pseudopalabras mostraron evidencias psicométricas satisfactorias de fiabilidad y validez interna, lo que apoya el establecimiento de normas estandarizadas para los años escolares de 2º a 5º año escolar.

Palabras clave: problemas de aprendizaje, dislexia, estudio de validación, lectura, psicomетría.

Artigo recebido: 10/09/2021; Artigo aceito: 22/08/2022.

Correspondências relacionadas a esse artigo devem ser enviadas a Douglas de Araújo Vilhena, Universidade Federal de Minas Gerais, Faculdade de Filosofia e Ciências Humanas – Av. Pres. Antônio Carlos, 6627 - Pampulha, Belo Horizonte – Minas Gerais, MG – Brasil, CEP 31270-901.

E-mail: douglasvilhena@ufmg.br

DOI:10.5579/rnl.2022.0742

Résumé

La présente recherche visait à fournir les sources de preuves psychométriques pour le "Test de reconnaissance de mots" et le "Test de reconnaissance de pseudo-mots" : (a) la fiabilité ; (b) la validité de la structure interne ; et (c) la normativité. Les participants étaient 598 élèves, 52% de sexe masculin, de la 2^{ème} à la 5^{ème} année de l'école élémentaire, âgés de 7 à 11 ans, stratifiés dans 102 classes de huit écoles publiques. Les scores des deux instruments sont donnés en termes de précision (pourcentages de mots lus correctement) et de taux de précision (nombre de mots lus correctement par minute). Comme preuve de fiabilité, il y avait une cohérence interne élevée et une équivalence entre les deux moitiés et les formes parallèles. La Validité de la Structure Interne a été vérifiée avec l'effet de la scolarité, la variable Taux de Précision montrant de meilleurs indices psychométriques que la Précision, avec une courbe de distribution normale et aucun chevauchement des Intervalles de Confiance entre les années scolaires. Il a été conclu que le test de reconnaissance de mots et le test de reconnaissance de pseudo-mots ont montré des preuves psychométriques satisfaisantes de fiabilité et de validité interne, ce qui soutient l'établissement de normes standardisées pour les années scolaires 2 à 5.

Mots clés : troubles d'apprentissage, dyslexie, étude de validation, lecture, psychométrie.

Abstract

The present research aimed to provide psychometric evidence for the 'Word Recognition Test' and 'Pseudoword Recognition Test' in relation to: (a) reliability; (b) validity of the internal structure; and (c) normatization. Participated 598 students, 52% male, from the 2nd to the 5th year of Elementary Education, 7 to 11 years old, stratified in 102 classrooms of eight state schools. The scores for both instruments are given terms of Accuracy (percentage of words read correctly) and Accuracy Rate (number of words read correctly per minute). Reliability evidence were the high internal consistency, split-half analyses and equivalence between parallel forms. The validity of the internal structure was verified with the effect of schooling, with the Accuracy Rate showing better psychometric indices than Accuracy, with normal distribution curve and without overlapping the Confidence Intervals between school years. In conclusion, the Word Recognition Test and the Pseudoword Recognition Test demonstrated satisfactory psychometric evidence of reliability and internal validity, which supported the establishment of standardized norms for the 2nd to the 5th school year.

Key words: dyslexia, learning disabilities, psychometrics, reading, validation study.

1. INTRODUÇÃO

As avaliações psicológicas e neuropsicológicas devem ser realizadas por meio de testes que apresentem características psicométricas satisfatórias, com diferentes fontes de evidências de fidedignidade e de validade, para assegurar a legitimidade do estabelecimento de normas e das interpretações dos resultados (Anuniação et al., 2020). Os instrumentos de avaliação da habilidade cognitiva de leitura que avaliam o reconhecimento visual de palavras, objetos do presente artigo, devem ser elaborados de acordo com o controle das variáveis psicolinguísticas recomendadas pela literatura, tais como frequência de ocorrência (Forster & Chambers, 1973), regularidade da relação grafema–fonema (Paap & Noel, 1991), número de letras (Rastle & Coltheart, 1998) e lexicalidade (Rastle & Coltheart, 1999). Ademais, devem embasar-se nos pressupostos dos modelos cognitivos-computacionais de leitura mais influentes: o modelo de dupla rota em cascata (Coltheart et al., 2001) e o modelo de processamento paralelo distribuído (Plaut et al., 1996).

Independentemente dos impasses existentes do ponto de vista das explicações com base em inferências e modelagem computacional, adotadas pela Ciência Cognitiva Experimental, na perspectiva das Neurociências, os estudos de neuroimagem, que exploram o contraste entre palavras com diferentes níveis de frequência, regularidade e comprimento, além do contraste entre palavras e pseudopalavras, aumentaram exponencialmente a compreensão dos mecanismos neurais durante o reconhecimento visual de palavras (Schuster et al., 2016), mostrando, portanto, a validade e realidade psicológica para os fenômenos descritos.

Outra preocupação ao se elaborar os testes de reconhecimento visual de palavras se refere à escolha da medida de desempenho, sendo as possibilidades a medida da 'acurácia' e a do 'tempo de processamento'. A acurácia, também referida como exatidão e precisão, é calculada como a porcentagem de palavras lidas corretamente, sendo a

variável mais utilizada nos instrumentos correntes. Os instrumentos que medem apenas a acurácia estão sujeitos ao efeito de teto, conforme frequentemente se observa (Pinheiro & Vilhena, 2022). Isso se deve ao português possuir estrutura silábica simples e transparência na correspondência grafema–fonema (Duncan et al., 2013; Pinheiro, 2011; Scliar-Cabral, 2003; 2019), o que propicia altos índices de acertos na leitura em voz alta de palavras e de pseudopalavras (Pinheiro & Vilhena, 2022). Já o tempo de processamento pode ser medido de duas formas: (a) 'tempo de reação', que é o período gasto entre o aparecimento do estímulo na tela de um computador e a emissão da resposta pelo participante; ou (b) 'taxa de acurácia', o computo do número de palavras lidas corretamente em um minuto.

A utilização do tempo de processamento, em adição à acurácia como escores de um teste, faz especial sentido nas ortografias transparentes (como a italiana e a alemã) em que o mapeamento entre os grafemas e fonemas, em comparação às ortografias profundas (como a inglesa e francesa), é mais regular e previsível. Isso porque no primeiro tipo de ortografia, em contraste com o segundo, a leitura nos estágios iniciais de aprendizagem (Marinelli et al., 2016), assim como as dificuldades nessa habilidade caracterizam-se por lentidão no reconhecimento de palavras, mesmo se acompanhado por alto índice de precisão (Wimmer, 1993; de Jong & van der Leij, 2003). Além disso, independentemente do tipo de ortografia, a compreensão de leitura pelo leitor em desenvolvimento é dependente do acesso preciso e, sobretudo, rápido ao léxico mental (Fuchs et al., 2001; Perfetti et al., 2005).

Portanto, é fundamental que a avaliação da habilidade cognitiva de leitura seja feita por meio de instrumentos que mensurem tanto a acurácia quanto uma ou outra forma de tempo de processamento, dependendo do tipo de teste. Para suprir essa lacuna no Brasil, o *Teste de Reconhecimento de Palavras* (TRP) e do *Teste de Reconhecimento de Pseudopalavras* (TRPp), de apresentação em cartões, foram elaborados para avaliar a leitura em voz

2. MÉTODO

2.1 Participantes

alta de crianças do segundo ao quinto ano do Ensino Fundamental (Pinheiro, 2013). Ambos os instrumentos apresentam a acurácia e a taxa de acurácia como escores. Fundamentam-se a partir de compreensivo referencial teórico sobre o reconhecimento visual de palavras, com fontes de validade de conteúdo satisfatórias, transcrições fonêmicas de cada item, assim como os efeitos psicolinguísticos de frequência, regularidade, extensão e lexicalidade (Pinheiro & Vilhena, 2022).

Como evidências de validade externa do TRP e TRPp, diferentes validades convergentes foram demonstradas pelas fortes correlações entre si e de moderadas a fortes com diferentes instrumentos de avaliação da leitura; validade discriminante com variáveis de dificuldades de saúde mental; e cinco diferentes validades de critério para prever o desempenho de leitura em voz alta de crianças (Pinheiro & Vilhena, 2022).

Embora o TRP e o TRPp já possuam fontes satisfatórias de validade de conteúdo e externa, é obrigatório que ambos os instrumentos demonstrem evidências de confiabilidade dos escores e da validade para avaliação do construto. Assim, a presente pesquisa objetivou prover as fontes de evidências psicométricas para o TRP e o TRPp no que concerne a: (a) fidedignidade (consistência interna, duas metades e formas paralelas); (b) validade da estrutura interna; e (c) normatização para o segundo ao quinto ano do Ensino Fundamental. Uma vez que o português brasileiro, diferentemente do português europeu, apresenta-se como uma ortografia transparente (Pinheiro, 2011; Scliar-Cabral, 2003, 2019) a hipótese é que a variável Taxa de Acurácia possua melhores índices psicométricos para diferenciar o desempenho de leitura do que a variável Acurácia. Caso os instrumentos demonstrem ser fidedignos e válidos para avaliar o desempenho de leitura, será possível o estabelecimento de normas representativas da população escolar, contribuindo com a avaliação nas áreas da Psicologia escolar e educacional no Brasil.

Todos os participantes forneceram consentimento informado e o Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais aprovou todos os procedimentos do estudo (número de identificação CAAE: 17754514.6.0000.5149). Respeitou-se todos os aspectos contidos na Resolução 466 de 2012 do Conselho Nacional de Saúde para pesquisa envolvendo seres humanos (Ministério da Saúde, 2013).

Participaram da pesquisa 598 alunos voluntários, 52,3% do sexo masculino, de 7 a 11 anos ($M = 9,6 \pm 1,1$ anos), do 2º ao 5º ano do Ensino Fundamental, todos falantes nativos do Português Brasileiro (ver Tabela 1). Conforme a listagem de todas as instituições registradas na cidade de Belo Horizonte, Minas Gerais, fornecida pela Superintendência Regional de Ensino, foram sorteadas oito escolas estaduais, estratificadas em diferentes regionais da cidade, não havendo recusa das instituições em participar da pesquisa. Em cada uma das 102 salas de aula, seis alunos foram selecionados, de forma aleatória (sorteio a partir da lista de presença). Detalhes adicionais sobre os participantes e sobre a bateria de avaliação cognitiva a que foram submetidos, incluindo a descrição dos Instrumentos e Procedimentos, encontram-se em diferentes estudos (Pinheiro et al., 2017; Vilhena et al., 2016; Vilhena & Pinheiro, 2016, 2020).

O tamanho amostral necessário foi estimado levando-se em consideração os seguintes parâmetros: erro de tolerância de $\pm 5\%$, Intervalo de Confiança de 95,0%, proporção populacional de 0,5 e população alvo de 157.875 crianças matriculadas no ensino fundamental de Belo Horizonte. O tamanho estatístico sugerido para uma amostra aleatória estratificada foi de 384 crianças, sendo acrescidas a esse valor 214 (46,7%) crianças, para formar a amostra final e garantir mais poder às análises.

Tabela 1.

Análise descritiva e distribuição da amostra (N = 598 participantes) de acordo com a idade, ano escolar, sexo e Percentil no Raven

Idade (anos)	Ano escolar				Total
	2º	3º	4º	5º	
7	67	0	0	0	67
8	121	71	1	0	193
9	0	62	70	3	135
10	0	3	59	82	144
11	0	0	0	59	59
Tamanho amostra (n)	188	136	130	144	598
Sexo (Masculino/Feminino)	102/86	73/63	65/65	73/71	313/285
Sexo Masculino (%)	54,3	53,7	50,0	50,7	52,3
Idade em anos (M ± DP)	8,1 ± 0,3	9,0 ± 0,4	9,9 ± 0,3	10,8 ± 0,4	9,6 ± 1,1
Idade em meses (M ± DP)	97,1 ± 3,9	107,4 ± 5,1	118,3 ± 3,6	130,1 ± 4,5	115,1 ± 13,0
Raven Percentil (M ± DP)	70,6 ± 23,7	79,0 ± 20,1	76,9 ± 19,1	71,9 ± 23,9	74,6 ± 22,0

2.2 Instrumentos

O Teste de Reconhecimento de Palavras (TRP) avalia a leitura de palavras isoladas, sendo composto por 88 palavras (4 de treinamento e 84 de teste) (Pinheiro, 2013). Os itens foram gerados a partir de um banco palavras aplicado

em uma amostra representativa de crianças e analisado via Teoria de Resposta ao Item, que considerou não apenas o acerto ou erro do aluno, mas também o tempo de reação gasto na leitura correta de cada palavra. As palavras selecionadas apresentaram índices de discriminação de médio a alto e três níveis de dificuldade (baixa, média e alta).

As palavras do TRP variaram em termos de sua frequência de ocorrência, sendo 40 classificadas como de alta frequência (no mínimo 130 ocorrências por milhão) (ex., *animais, caderno, terra*) e 44 de baixa frequência (de 6 a 47 ocorrências por milhão) (ex., *atleta, monarca, sacola*). Essa classificação se baseou no banco de dados da contagem de frequência de ocorrência de palavras no português brasileiro (Pinheiro, 2015). Além do controle de frequência, as palavras do TRP foram classificadas em dois níveis de regularidade de acordo com a decodificação de cada grafema que as constitui [34 semi-regulares (SR) e 50 irregulares (IR)] e em três níveis de extensão [37 palavras curtas (quatro ou cinco letras), 26 palavras médias (seis letras) e 21 palavras longas (sete ou oito letras)].

O Teste de Reconhecimento de Pseudopalavras (TRPp) avalia a leitura de pseudopalavras isoladas, sendo composto por 88 itens (4 de treino e 84 de teste) (Pinheiro, 2013). O instrumento foi construído a partir do TRP, sendo que cada pseudopalavra mantém a mesma estrutura ortográfica e número de letras da palavra de referência (ex., *pederno, clandas, verra*).

O TRP e o TRPp, ao mesmo tempo que testam construtos diferentes (processamento lexical versus decodificação, respectivamente), se completam, portanto, devem ser aplicados juntos, um após o outro. Por exemplo, o efeito lexical ou lexicalidade só pode ser averiguado a partir da comparação dos resultados da leitura de palavras e de pseudopalavras (Rastle & Coltheart, 1999). Quanto a ordem de aplicação, em situação de clínica aplica-se TRP, em primeiro lugar, seguido do TRPp. Em situações experimentais, para o controle do 'efeito de ordem', a ordem de aplicação dos instrumentos é intercalada entre os participantes.

O escore da variável Acurácia, para fins de comparação entre instrumentos com números de itens diferentes, é dado em porcentagem. Ou seja, o número total de palavras ou pseudopalavras lidas corretamente (máximo de 84 pontos), dividido por 84, vezes cem (ex., 75 palavras corretas correspondem à Acurácia de 89%). Já o escore da variável Taxa de Acurácia consiste do número total de palavras ou pseudopalavras lidas corretamente (máximo de 84 pontos) vezes sessenta, dividido pelo tempo total em segundos (ex., 75 palavras corretas em 120 segundos correspondem a Taxa de Acurácia de 37,5 palavras por minuto).

2.3 Procedimentos

Cada participante foi individualmente submetido ao TRP e o TRPp (administrados em sequência, mas em ordem aleatória). O TRP e TRPp são apresentados em cartões plastificados, cada um contendo 11 linhas, com seis a nove palavras por linha, impressos em tinta preta em papel branco A4, em fonte Arial, tamanho 14. Para prover fonte de evidência de fidedignidade e para evitar o efeito de ordem, cada instrumento possui três formas paralelas, cujas versões variam apenas em relação a ordem das palavras (Versão A, B e C) e das pseudopalavras (Versão D, E e F).

Os instrumentos foram administrados por um psicólogo e seis estudantes de graduação em Psicologia treinados para a aplicação. Foi requerido aos participantes,

individualmente, que lessem em voz alta o TRP e o TRPp (apresentados de forma randômica) com o maior número de acerto e o mais rápido possível. Os instrumentos são fáceis de serem administrados e de pontuar. Cada um inicia com quatro itens de treino, seguidos dos 84 itens de teste, lidos pelo participante da esquerda para a direita, de cima para baixo. Nos crivos de correção de cada teste (papel e lápis), os aplicadores registraram os erros de acurácia de leitura de forma dicotômica (certo e errado) e o tempo total de leitura em segundos. A avaliação foi interrompida após 10 erros consecutivos. Considerou-se acurada apenas a leitura fluente para cada item. Por fim, destaca-se que a autocorreção (item lido incorretamente, mas imediatamente corrigido) foi considerada como erro.

2.4 Análise de dados

As análises estatísticas foram realizadas por meio do software *IBM SPSS Statistics* (versão 21.0, Chicago, Illinois, EUA). Não se encontraram valores atípicos em nenhum instrumento ao utilizar a técnica *Outlier Labeling Rule*, com valor de g de 2,2. Para verificar a distribuição dos dados, os valores de assimetria e de curtose foram divididos pelos seus respectivos erros padrões, sendo significantes os valores acima de 1,96 (valores em negrito na Tabela 2).

A distribuição foi confirmada pelos testes de normalidade *Kolmogorov-Smirnov* e de *Shapiro-Wilk*. Uma vez que a variável Acurácia não apresentou uma distribuição normal, foram utilizados os testes de análise de variância não-paramétricos para a comparação de sexo (*U* de *Mann-Whitney*) e para a análise do efeito de escolaridade (*H* de *Kruskal-Wallis*). A variável Taxa de Acurácia apresentou distribuição normal, sendo utilizado o teste paramétrico de Análise de Variância (ANOVA). A consistência interna foi estimada pelo coeficiente *Kuder-Richardson* [fórmula 20 (*KR-20*), específico para variáveis dicotômicas (certo e errado)] e pelo método de duas metades (*Spearman-Brown* e *Guttman*). Significância estatística foi estabelecida em $p < 0,05$.

3. RESULTADOS

3.1 Evidências de fidedignidade

Como evidência de fidedignidade, foram verificados os coeficientes de consistência interna (*Kuder-Richardson*, *Spearman-Brown* e *Guttman*) e a equivalência (formas paralelas). Com relação à consistência interna, ao se analisar em conjunto todos os itens de teste do TRP e do TRPp ($N = 168$ estímulos), foi possível verificar grau de covariância próximo ao máximo ($KR-20 = 0,98$). O TRP ($n = 84$ itens) apresentou consistência interna muito alta ($KR-20 = 0,96$), inclusive ao se analisar separadamente as palavras de alta frequência (0,93), baixa frequência (0,92), semi-regulares (0,91), irregulares (0,93), tamanho curto (0,92), médio (0,87), e longo (0,85). Da mesma forma, o TRPp ($n = 84$ itens) apresentou consistência interna muito alta ($KR-20 = 0,96$), mesmo ao analisar apenas as pseudopalavras curtas (0,91), médias (0,88), e longas (0,87).

Na análise da consistência interna pelo método das duas metades (*split-half*), os itens foram repartidos em duas partes (Pares vs. Ímpares), tendo apresentado índices de correlações *Spearman-Brown* e *Guttman* muito altos para o TRP (0,95) e para o TRPp (0,96). Como evidência de equivalência, não houve diferença entre as três versões paralelas utilizadas do TRP [Versão A, Versão B e Versão C ($F_{(2, 478)} = 1,7, p = 0,19$)] e as do TRPp [Versão D, Versão E e Versão F ($F_{(2, 478)} = 1,1, p = 0,33$)].

3.2 Evidências de validade da estrutura interna

A variável Acurácia demonstrou limitações nas evidências de validade com base na estrutura interna. A distribuição amostral apresentou assimetria negativa e uma curtose leptocúrtica, tanto para a amostra geral ($N = 598$) quanto por ano escolar (ver valores em negrito na Tabela 2), ambas confirmadas como significantes ($p < 0,001$) nos testes de *Kolmogorov-Smirnov* e de *Shapiro-Wilk*.

A Acurácia apresentou significativo efeito de escolaridade para o TRP [$H_{(3, 594)} = 53,5, p < 0,001$] e para o TRPp [$H_{(3, 594)} = 32,5, p < 0,001$]: na análise *Post Hoc*, foi confirmada a diferença significativa entre os anos escolares (2°

< 3° < 4° < 5° ano). Os Intervalos de Confiança de 95% da Acurácia se sobrepueram tanto no TRP (2° ano: 83,3–89,8; 3° ano: 89,4–93,4; 4° ano: 92,6–95,0; 5° ano: 94,6–96,5) quanto no TRPp (2° ano: 73,3–80,3; 3° ano: 79,4–85,0; 4° ano: 80,1–84,6; 5° ano: 85,0–88,5).

A variável Taxa de Acurácia demonstrou melhores índices de validade da estrutura interna (Tabela 2). A Taxa de Acurácia apresentou uma distribuição normal padrão, sendo simétrica e mesocúrtica, tanto para o TRP (Figura 1) quanto para o TRPp (Figura 2). Essa normalidade foi confirmada pelos resultados não significantes ($p > 0,31$) dos testes de *Kolmogorov-Smirnov* e de *Shapiro-Wilk* (exceto a assimetria positiva no 2° ano para o TRP, assinalada em negrito na Tabela 2).

A Taxa de Acurácia apresentou significativo efeito de escolaridade para o TRP [$F_{(3, 594)} = 63,2, MSE = 391,4, p < 0,001$] e para o TRPp [$F_{(4, 594)} = 47,3, MSE = 129,2, p < 0,001$]. Os Intervalos de Confiança de 95% da Taxa de Acurácia não se sobrepueram, tanto para o TRP (2° ano: 27,3–33,8; 3° ano: 39,6–47,3; 4° ano: 51,2–58,3; 5° ano: 60,9–67,9), quanto para o TRPp (2° ano: 17,9–21,5; 3° ano: 23,8–28,2; 4° ano: 28,4–32,3; 5° ano: 34,8–38,9).

Tabela 2.

Análise descritiva das variáveis Acurácia (%) e Taxa de Acurácia do TRP e do TRPp, por ano escolar

Análise descritiva	TRP					TRPp					
	Geral	2° ano	3° ano	4° ano	5° ano	Geral	2° ano	3° ano	4° ano	5° ano	
Acurácia	Média (%)	92,4	86,5	91,7	93,8	95,5	82,5	76,8	82,2	82,3	86,7
	Desvio Padrão	10,7	16,2	11,6	6,8	5,6	14,1	17,2	14,5	12,9	10,7
	Assimetria/Erro	-3,7/1	-2,5/2	-3,4/2	-2,7/2	-2,6/2	-1,8/1	-1,7/3	-1,7/2	-1,3/2	-1,6/2
	Curtose/Erro	17,7/2	6,9/5	15,3/5	10,5/4	8,8/4	4,0/2	3,4/5	3,2/5	2,2/4	2,7/4
	Mínimo	17,0	17,0	20,0	54,5	64,8	11,4	11,4	20,0	29,5	45,1
	Máximo	100	100	100	100	100	100	98,9	98,8	97,7	100
Taxa de Acurácia	Média (ppm)	50,2	30,6	43,4	54,7	64,4	29,2	19,7	26,1	30,3	36,8
	Desvio Padrão	23,3	16,3	19,9	20,5	21,1	12,9	9,1	11,5	11,5	12,5
	Assimetria/Erro	0,1/1	0,7/0,2	0,2/0,2	-,1/2	-0,1/2	0,4/1	0,3/0,2	0,4/0,2	0,4/0,2	0,2/2
	Curtose/Erro	-0,6/2	0,7/0,5	-0,2/5	-,5/4	-0,5/4	-0,1/2	0,2/0,5	0,1/0,5	-0,3/4	0,0/4
	Mínimo	1,0	1,0	1,8	9,6	7,2	1,9	2,0	1,9	5,8	4,3
	Máximo	110,0	83,2	93,2	101,2	110,0	71,8	48,3	57,2	60,0	71,8

Nota. Valores em negrito são estatisticamente significantes. %: porcentagem de palavras lidas corretamente; ppm: número de palavras lidas corretamente por minuto.

3.3 Estudo de normatização

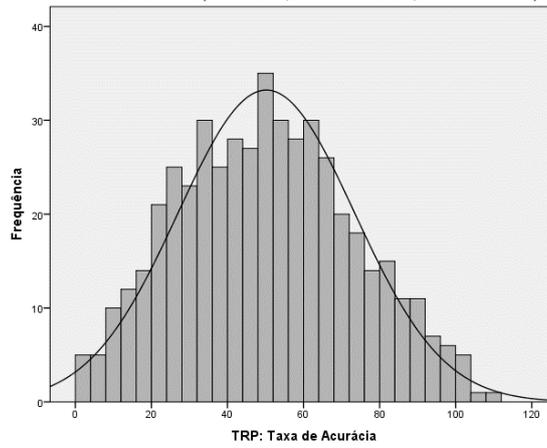
Uma vez que o Teste de Reconhecimento de Palavras e o Teste de Reconhecimento de Pseudopalavras demonstraram fontes de evidências psicométricas satisfatórias, foi possível sustentar o estabelecimento de normas padronizadas para o 2° ao 5° ano ($N = 598$), com as variáveis Acurácia e Taxa de Acurácia convertidas para o Percentil correspondente (Tabela 3). Foram criados pontos de

corte para a classificação do desempenho de leitura entre transtorno (Percentil 0 a 7), abaixo da média (Percentil 8 a 25), média (Percentil 26 a 74), acima da média (Percentil 75 a 94) e alto (Percentil 95 a 99). Uma análise de variância univariada (ANOVA), com uma correção de Bonferroni, confirmou diferenças significantes nas pontuações para todas as cinco classificações de desempenho de leitura ($p < 0,001$).

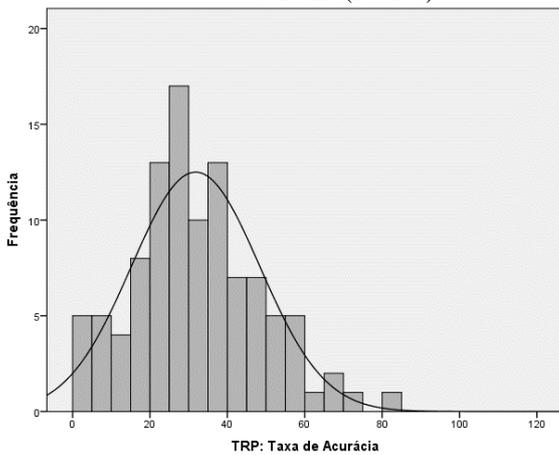
Figura 1.

Frequência (número de participantes) por escore da Taxa de Acurácia do Teste de Reconhecimento de Palavras para a amostra total (Figura 1.a) e para cada ano escolar (Figura 1.b – f), ajustada com uma curva de distribuição normal

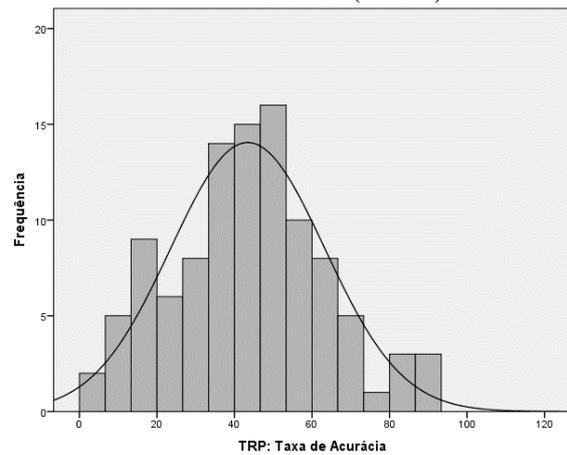
a. Amostra total (N = 598; 2º ao 5º ano; 7 a 11 anos)



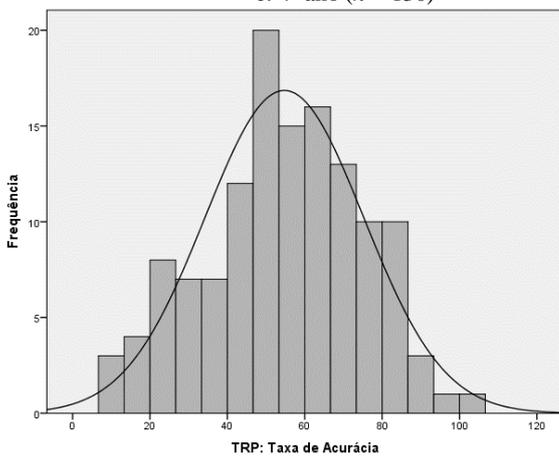
c. 2º ano (n = 188)



d. 3º ano (n = 136)



e. 4º ano (n = 130)



f. 5º ano (n = 144)

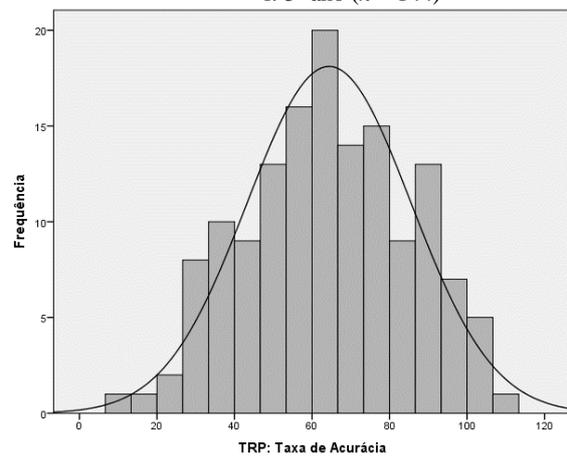
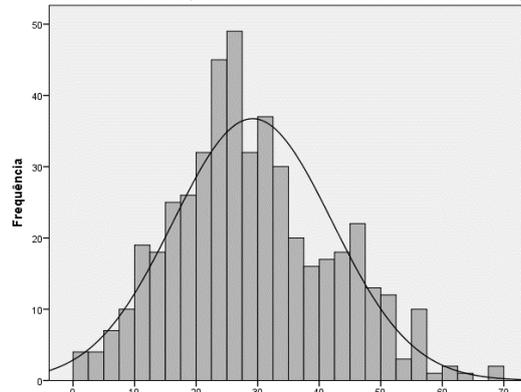


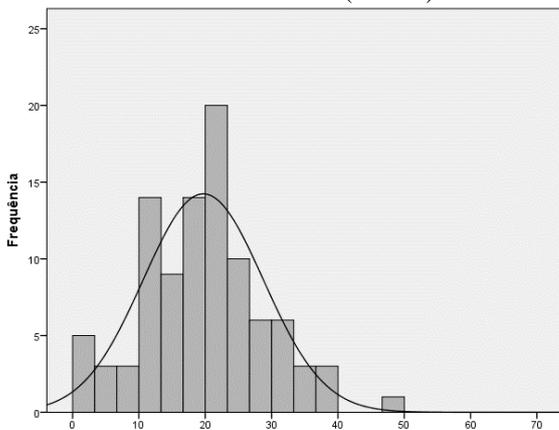
Figura 2.

Frequência (número de participantes) por escore da Taxa de Acurácia do Teste de Reconhecimento de Pseudopalavras para a amostra total (Figura 2.a) e para cada ano escolar (Figura 2.b–f), ajustada com uma curva de distribuição normal

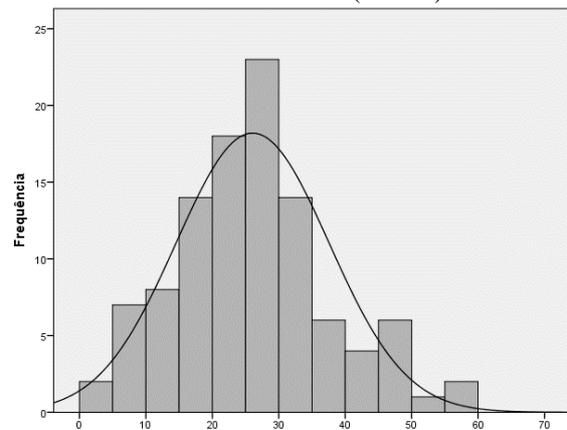
a. Amostra total (N = 598; 2º ao 5º ano; 7 a 11 anos)



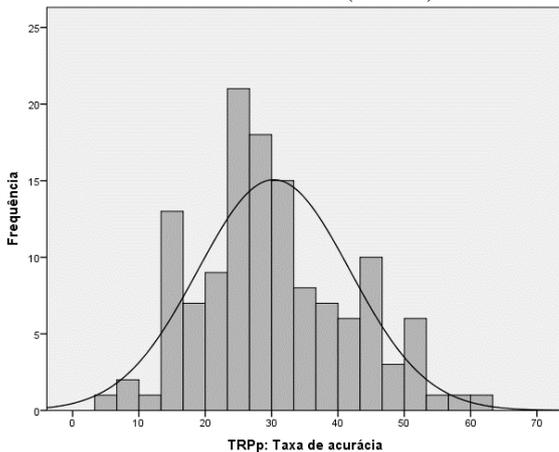
c. 2º ano (n = 188)



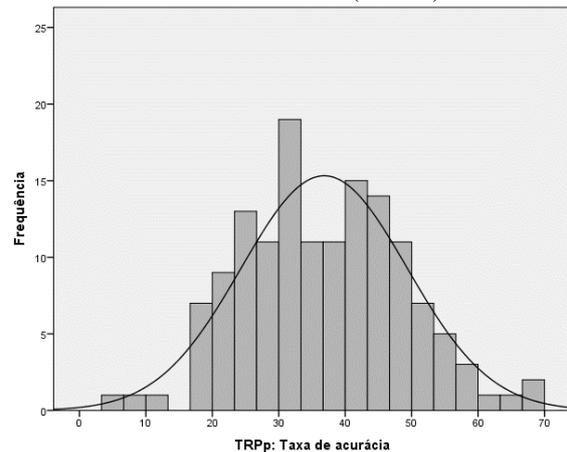
d. 3º ano (n = 136)



e. 4º ano (n = 130)



f. 5º ano (n = 144)



4. DISCUSSÃO

A presente pesquisa objetivou prover as fontes de evidências psicométricas para o TRP e o TRPp no que concerne a fidedignidade, validade da estrutura interna e normatização. Ambos os instrumentos alcançaram evidências psicométricas satisfatórias, em especial para a variável Taxa de Acurácia, o que sustentou o estabelecimento de normas padronizadas para os alunos do segundo ao quinto ano do Ensino Fundamental. Estes resultados complementam os

dados encontrados em estudos prévios (Cogo-Moreira et al., 2012; Pinheiro, 2013; Pinheiro et al., 2017; Pinheiro & Vilhena, 2022; Vilhena et al., 2016; Vilhena & Pinheiro, 2016; 2020).

Foram verificadas diferentes fontes de evidências de fidedignidade (consistência interna, duas metades e formas paralelas). Os resultados de consistência interna pelo coeficiente *Kuder-Richardson* evidenciaram grau de covariância próximo ao máximo, com evidência forte de que os itens do TRP e TRPp medem o mesmo construto (i.e.,

decodificação grafêmica). Os resultados das análises pelo método das duas metades demonstraram que o mesmo indivíduo obteve um escore médio equivalente nas duas metades do TRP e do TRPp, o que demonstra a homogeneidade dos instrumentos. Houve equivalência entre

as três formas paralelas do TRP e as três do TRPp, o que permite alternar, sem prejuízo psicométrico, a apresentação das versões para evitar o efeito de ordem dos itens.

Tabela 3.

Estudo de Normatização – Conversão das variáveis Acurácia e da Taxa de Acurácia, do TRP e do TRPp, para Percentil e respectiva classificação de desempenho de leitura de acordo com o ano escolar do participante

	Desempenho em leitura	Percentil	TRP				TRPp			
			2º ano	3º ano	4º ano	5º ano	2º ano	3º ano	4º ano	5º ano
Acurácia (%)	<i>Transtorno</i>	7	60	74	85	87	49	55	59	66
	<i>Abaixo da média</i>	25	85	91	92	94	71	77	77	84
	<i>Média</i>	50	92	96	96	98	80	86	84	90
	<i>Acima da média</i>	75	96	98	99	99	89	93	93	94
	<i>Alto</i>	95	99	100	100	100	94	97	98	99
Taxa de Acurácia (ppm)	<i>Transtorno</i>	7	8	13	21	31	6	9	15	19
	<i>Abaixo da média</i>	25	21	32	41	49	13	18	23	27
	<i>Média</i>	50	28	43	55	65	20	26	29	36
	<i>Acima da média</i>	75	40	56	70	80	25	33	38	46
	<i>Alto</i>	95	63	81	86	99	36	49	51	57

Nota. %: porcentagem de palavras lidas corretamente; ppm: número de palavras lidas corretamente por minuto.

Em relação às evidências de validade baseada na estrutura interna, tanto a variável Acurácia quanto a Taxa de Acurácia conseguiram distinguir os leitores por ano de escolaridade (2º < 3º < 4º < 5º ano). Esse resultado está de acordo com o estudo de Cogo-Moreira et al. (2012), que investigou uma amostra estratificada de 617 alunos ($M = 9,8 \pm 1,0$ anos), por meio da análise de regressão de *Tobit* ajustada para os *clusters* de 10 escolas, tendo demonstrado que a Acurácia tanto do TRP (3º ano: $\beta = 6,62, p < 0,01$; 4º ano $\beta = 10,56, p < 0,01$) quanto do TRPp (3º ano $\beta = 4,45, p < 0,001$; 4º ano $\beta = 6,77, p < 0,001$) possuem efeito de escolaridade significativa, corroborando as evidências de validade da estrutura interna.

No entanto, encontrou-se, como esperado, que a variável Taxa de Acurácia apresenta melhores índices de validade de estrutura interna do que a variável Acurácia, tanto para o TRP quanto para o TRPp. A Taxa de Acurácia possui uma curva padrão de distribuição normal (Figura 1 e Figura 2), sendo simétrica (implica que a média, a mediana e a moda são coincidentes) e mesocúrtica (grau de achatamento normal), sem sobreposição dos Intervalos de Confiança entre os anos escolares. Estes resultados revelam a importância de as avaliações da habilidade de leitura de palavras e pseudopalavras considerarem não somente a Acurácia, mas também o tempo necessário para finalizar o teste.

As normatizações dos escores do TRP e do TRPp foram apresentadas, uma vez que os instrumentos mostraram fontes de evidências de fidedignidade e validade satisfatórias para alunos do 2º ao 5º ano do ensino fundamental. Uma força do estudo é a amostragem aleatória, estratificada e com

tamanho estatístico adequado. Nesse sentido, foi possível estabelecer normas de conversão das variáveis Acurácia e da Taxa de Acurácia para o seu correspondente Percentil e classificação do desempenho de leitura.

O Manual Diagnóstico e Estatístico-5 (DSM-5) (APA, 2013) recomenda uma pontuação de corte no Percentil 7 para um diagnóstico de Transtorno Específico de Aprendizagem (neste caso, com o especificador de prejuízo na leitura), o que evita falsos positivos quando utilizado para fins clínicos (maior especificidade). O ponto de corte no Percentil 25 indica uma habilidade de leitura bem abaixo da média esperada para o ano escolar, sendo um preditor de dificuldade de leitura, especialmente para fins de pesquisa (maior sensibilidade).

Finalmente, o Teste de Reconhecimento de Palavras e o Teste de Reconhecimento de Pseudopalavras se primam devido: (a) a sua concepção teórica, que permite o teste de efeitos psicolinguísticos clássicos encontrados na literatura e replicados tanto nas ortografias opacas (Provazza et al. 2019; Schuster et al., 2016) quanto nas ortografias transparentes (Carreiras et al., 2006; Pinheiro, 1995; 1999; Wimmer et al., 2010); (b) à utilização de escores de acurácia e de taxa de acurácia como medida de desempenho, que em conjunto possibilitam discriminação da proficiência dos leitores; e (c) às suas propriedades psicométricas, que lhes conferem fidedignidade, validade e normas. Essas características tornam esses instrumentos úteis para a avaliação cognitiva da leitura tanto no contexto clínico quanto na pesquisa.

5. CONCLUSÃO

O Teste de Reconhecimento de Palavras e o Teste de Reconhecimento de Pseudopalavras demonstraram fontes de evidências psicométricas de fidedignidade (equivalência e consistência interna) e de validade da estrutura interna satisfatórias para a avaliação das habilidades de leitura de crianças de escolas públicas cursando do 2º ao 5º ano do ensino fundamental. O estudo de normatização populacional apresentou como vantagens a amostragem aleatória, estratificada e com tamanho estatístico adequado. É importante ressaltar que apenas após a demonstração das evidências de fidedignidade e de validade e da robustez teórica de um teste é que se torna possível estabelecer normas populacionais e clínicas para caracterizar os leitores. No que se refere ao TRP e ao TRPp, enfatiza-se ainda a importância de que se desenvolvam normas representativas para diferentes tipos de escolas (municipais e particulares) e em diferentes regiões do Brasil.

Referências

- Anunciação, L., Portugal, A. C. A., Rabelo, I., Cruz, R., & Landeira-Fernandez J. (2020). Propriedades psicométricas de instrumento de Memória Visual de Curto Prazo (MEMORE). *Neuropsicologia Latinoamericana*, 12(2), 44-58. https://www.neuropsicolatina.org/index.php/Neuropsicologia_Latinoamericana/article/view/545
- APA (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). American Psychiatric Association.
- Carreiras, M., Mechelli, A., & Price, C. J. (2006). Effect of word and syllable frequency on activation during lexical decision and reading aloud. *Hum Brain Mapp*, 27(12), 963-972. <https://doi.org/10.1002/hbm.20236>
- Cogo-Moreira, H., Ploubidis, G., De Avila, C., Mari, J., & Pinheiro, A. M. V. (2012). EACOL (Scale of Evaluation of Reading Competency by the Teacher): Evidence of concurrent and discriminant validity. *Neuropsychiatric Diseases and Treatment*, 8(1), 443-454. <https://doi.org/10.2147/NDT.S36196>
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204-256. <https://doi.org/10.1037/0033-295X.108.1.204>
- de Jong, P. F., & van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology*, 95(1), 22-40. <https://doi.org/10.1037/0022-0663.95.1.22>
- Duncan, L. G., Castro, S. L., Defior, S., Seymour, P. H., Baillie, S., Leybaert, J., Mousty, P., Genard, N., Sarris, M., Porpodas, C. D., Lund, R., Sigurdsson, B., Práinsdóttir, A. S., Sucena, A., & Serrano, F. (2013). Phonological development in relation to native language and literacy: variations on a theme in six alphabetic orthographies. *Cognition*, 127(3), 398-419. <https://dx.doi.org/10.1016/j.cognition.2013.02.009>
- Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, 12(6), 627-635. [https://doi.org/10.1016/S0022-5371\(73\)80042-8](https://doi.org/10.1016/S0022-5371(73)80042-8)
- Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: a theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, 5(3), 239-256. https://dx.doi.org/10.1207/S1532799XSSR0503_3
- Marinelli, C. V., Romani, C., Burani, C., McGowan, V. A., & Zoccolotti, P. (2016). Costs and benefits of orthographic inconsistency in reading: evidence from a cross-linguistic comparison. *PLoS One*, 11(6), e0157457. <https://dx.doi.org/10.1371/journal.pone.0157457>
- Ministério da Saúde (2013). Conselho Nacional de Saúde. Resolução 466, de 12 de dezembro de 2012. Diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos. *Diário Oficial da União [da] República Federativa do Brasil*, 150 (112 seção 1), 59-62.
- Paap, K. R., & Noel, R. W. (1991). Dual route models of print to sound: Still a good horse race. *Psychological Research*, 53(1), 13-24. <https://doi.org/10.1007/BF00867328>
- Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The Acquisition of Reading Comprehension Skill. In M. J. Snowling & C. Hulme (Eds.), *Blackwell handbooks of developmental psychology. The science of reading: A handbook* (p. 227-247). Blackwell Publishing. <http://dx.doi.org/10.1002/9780470757642.ch13>
- Pinheiro, A. M. V. (1995). Reading and spelling development in Brazilian Portuguese. *Reading & Writing*, 7(1), 111-138. <https://doi.org/10.1007/BF01026950>
- Pinheiro, A. M. V. (1999). Cognitive assessment of competent and impaired reading in Scottish and Brazilian children. *Reading & Writing*, 11, 175-211. <https://doi.org/10.1023/A:1007914412313>
- Pinheiro, A. M. V. (2011). Transparência ortográfica e o efeito de retroalimentação fonológico grafêmica: implicações para a construção de provas de reconhecimento de palavras. In L. M. Alves, R. Mousinho, & S. A. Capellini (Eds.). *Dislexia: Novos temas, novas perspectivas* (pp. 131-146). Wak.
- Pinheiro, A. M. V. (2013). *Prova de Leitura e de Escrita de palavras e de pseudopalavras*. Relatório Técnico Final aprovado pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG. Número do processo: APQ-01914-09.
- Pinheiro, A. M. V. (2015). Frequency of Occurrence of Words in Textbooks Exposed to Brazilian children in the Early Years of Elementary School. Updated database from 1996. *CHILDES – Child Language Data Exchange System*. <http://childes.talkbank.org/derived>
- Pinheiro, A. M. V., & Vilhena, D. A. (2022). Teste de reconhecimento de palavras e pseudopalavras: validades de conteúdo e externa. *Signo*, 47(88), 147-164. <https://doi.org/10.17058/signo.v47i88.17396>
- Pinheiro, A. M. V., Vilhena, D. A., & Santos, M. A. C. (2017). PROLEC – Provas de Avaliação dos Processos de Leitura: análise de suas características psicométricas. *Trends in Psychology [Temas em Psicologia]*, 25(3), 1067-1080. <http://dx.doi.org/10.9788/tp2017.3-08>
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review*, 103(1), 56-115. <http://dx.doi.org/10.1037/0033-295X.103.1.56>
- Provazza, S., Giofrè, D., Adams, A. M., & Roberts, D. J. (2019). The Clock Counts - Length Effects in English Dyslexic Readers. *Frontiers in psychology*, 10, 2495. <https://doi.org/10.3389/fpsyg.2019.02495>
- Rastle, K., & Coltheart, M. (1998). Whammy and double whammy: The effect of length on nonword reading. *Psychonomic Bulletin and Review*, 5(2), 277-282. <http://dx.doi.org/10.3758/BF03212951>
- Rastle, K., & Coltheart, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology*:

- Human Perception and Performance*, 25(2), 482-503.
<https://doi.org/10.1037/0096-1523.25.2.482>
- Schuster, S., Hawelka, S., Hutzler, F., Kronbichler, M., & Richlan, F. (2016). Words in Context: The Effects of Length, Frequency, and Predictability on Brain Responses During Natural Reading. *Cerebral Cortex*, 26(10), 3889–3904.
<https://doi.org/10.1093/cercor/bhw184>
- Scliar-Cabral, L. (2003). *Princípios do Sistema Alfabético do Português do Brasil*. Editora Contexto.
- Scliar-Cabral, L. (2019). Precisamos de mais de 50.000 palavras no léxico ortográfico para ler? *Letras de Hoje*, 54(2), 122-131. <http://dx.doi.org/10.15448/1984-7726.2019.2.34510>
- Vilhena, D. A., & Pinheiro, A. M. V. (2016). Revised version of the Scale of Evaluation of Reading Competence by the Teacher: final validation and standardization. *Universitas Psychologica*, 15(4), 1-13.
<http://dx.doi.org/10.11144/Javeriana.upsy15-4.efvs>
- Vilhena, D. A., & Pinheiro, Â. M. V. (2020). Fidedignidade, Validade e Normatização do Teste de Leitura: Compreensão de Sentenças. *Psicologia: Teoria e Pesquisa*, 36, e36325.
<https://dx.doi.org/10.1590/0102.3772e36325>
- Vilhena, D. A., Sucena, A., Castro, S. L., & Pinheiro, Â. M. V. (2016). Reading Test-Sentence Comprehension: An Adapted Version of Lobrot's Lecture 3 Test for Brazilian Portuguese. *Dyslexia*, 22(1), 47-63.
<http://dx.doi.org/10.1002/dys.1521>
- Wimmer, H. (1993). Characteristics of developmental dyslexia in a regular writing system. *Applied Psycholinguistics*, 14(1), 1-33. <https://dx.doi.org/10.1017/S0142716400010122>
- Wimmer, H., Schurz, M., Sturm, D., Richlan, F., Klackl, J., Kronbichler, M., & Ladurner, G. (2010). A dual-route perspective on poor reading in a regular orthography: An fMRI study. *Cortex*, 46(10), 1284-1298.
<https://doi.org/10.1016/j.cortex.2010.06.004>