

Construção de uma Tarefa de Leitura em Voz Alta de Palavras: Análise Psicométrica dos Itens

The Construction of a Word Reading Aloud Task: A Psychometric Item Analysis Report

Patrícia Silva Lúcio^{*a}, Ricardo José de Moura^b, Elizabeth do Nascimento^b & Ângela Maria Vieira Pinheiro^b
^aUniversidade Estadual de Londrina, Londrina, Brasil & ^bUniversidade Federal de Minas Gerais, Belo Horizonte, Brasil

Resumo

O estudo descreve a análise dos itens que irão compor uma tarefa de leitura em voz alta de palavras isoladas. Um grupo de crianças do 2º ao 5º ano do Ensino Fundamental leu, por meio de microcomputadores, uma lista contendo 323 palavras de baixa frequência. Foram selecionados 112 itens-candidatos que apresentaram índices adequados de dificuldade, discriminação, correlação item-total e consistência interna. A análise mostrou que a maioria das palavras apresentou níveis de dificuldade baixos, mas paralelamente produziu níveis adequados de discriminação e de correlação item-total. Elevados índices de consistência interna foram obtidos, indicando a homogeneidade da tarefa e a unidimensionalidade do construto avaliado. Considerações de Embretson (1983) sobre as pesquisas de representação do construto e de intervalo nomotético são discutidas.

Palavras-chave: Leitura em voz alta, análise de itens, dificuldade, discriminação, consistência interna.

Abstract

The study reports the psychometric inquiry of a word reading aloud task. A group of 1st to 4th grade school children read aloud 323 low frequency words that were presented in a computer screen. The item analysis selected 112 potential items that presented satisfactory levels of difficulty, discrimination, item total correlation, and internal consistency. Generally, the selected items were not difficult to the sample. Nevertheless, most items had good levels of discrimination and item total correlation. The internal consistency index attested for the homogeneity of the task and for the construct unidimensionality. Considerations of Embretson (1983) about construct representation research and nomothetic span research are outlined.

Keywords: Word reading aloud, item analysis, difficulty, discrimination, internal consistency.

Em psicologia, a avaliação e o diagnóstico de competências psicológicas específicas, como a inteligência e as habilidades de leitura e de escrita, necessitam da utilização de instrumentos que apresentem evidências de validade, precisão e normas para a população a que se destinam. No Brasil, a carência de sistematização da maioria dos instrumentos de avaliação psicológica conduziu a uma resolução por parte do Conselho Federal de Psicologia (CFP, 2003) que define e regulamenta o uso, a elaboração e a comercialização de testes psicológicos. O artigo quatro dessa resolução é dedicado ao estabelecimento dos requisitos mínimos necessários para o reconhecimento oficial de instrumentos que buscam estabelecer uma avaliação

objetiva das propriedades psicológicas. Dentre outras recomendações, essa resolução ressalta a necessidade de especificação da fundamentação teórica do instrumento, com ênfase na definição do construto, dos propósitos e dos principais contextos para os quais ele foi desenvolvido. Determina também que os instrumentos devam passar pelos procedimentos de empíricos de investigação de suas propriedades psicométricas.

Após a iniciativa conduzida pelo CFP – que pode ser ela mesma considerada fruto de uma inquietação por parte da sociedade a partir da década de 1990 sobre a qualidade dos instrumentos de medida disponíveis no Brasil (Amendola, 2006) –, observou-se uma revitalização da pesquisa na área. Esta revitalização é representada pela criação de novos instrumentos de medida e atualização dos já existentes, além dos esforços de entidades como o Instituto Brasileiro de Avaliação Psicológica (IBAP) na promoção de encontros científicos e publicações de livros e artigos científicos sobre o tema (Noronha, Primi, & Alchieri, 2005).

No entanto, especificamente em relação à avaliação da habilidade de reconhecimento de palavras, essas inicia-

* Esta pesquisa é parte da dissertação de mestrado da primeira autora, orientada pela terceira e quarta autoras. Agradecemos ao apoio financeiro da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Endereço para correspondência: Rua Costa Rica, 455, Bairro Nações Unidas, Sabará, Belo Horizonte, MG, Brasil 34590-380. Tel.: (31) 3499-6268. E-mail: pslucio@gmail.com.

tivas ainda se mostram muito tímidas no contexto brasileiro, sendo raros os empreendimentos voltados para a construção de instrumentos que seguem as etapas propostas pela psicometria (p. ex.: Chadha, 2009; Pasquali, 1996; Urbina, 2004/2007). O que se verifica nos trabalhos publicados em revistas e livros é a avaliação dessa habilidade a partir do uso de listas experimentais de palavras, destinadas a examinar os componentes dos modelos cognitivos que são utilizados pelo leitor durante a leitura, como o modelo de dupla-rota (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Levy et al., 2009). Este empreendimento é conseguido por meio da observação de efeitos na leitura, obtidos a partir do controle das características psicolinguísticas das palavras, tais como a frequência e a regularidade. Pinheiro e Rothe-Neves (2001) fazem recomendações sobre como deve ser conduzido esse controle, ressaltando a necessidade de se estabelecer um pareamento de quantas características psicolinguísticas das palavras forem possíveis. Assim, o conjunto final de itens deve conter número semelhante de palavras familiares e não-familiares, regulares e irregulares, com maior ou menor número de sílabas ou letras e assim por diante.

O controle exclusivo das características psicolinguísticas das palavras, em detrimento da investigação empírica de suas propriedades psicométricas, implica que a validade do procedimento de construção de tarefas de leitura em voz alta tem sido feita exclusivamente via teoria¹. Em relação a este tipo de tarefa, parece haver apenas duas exceções em nosso contexto², a saber, os subtestes de leitura do Teste de Desempenho Escolar (Stein, 1994) e da bateria PROLEC, recentemente adaptada para o português pelo grupo de Capellini (Cuetos, Rodrigues, & Ruano, 1998).

O subteste de leitura do TDE apresenta normas para crianças da 1.^a à 6.^a série do Ensino Fundamental, além de evidências psicométricas de validade e fidedignidade. Apesar disso, o subteste não deixa de apresentar proble-

mas. Em um estudo recente, Lúcio, Pinheiro e Nascimento (2009) avaliaram o desempenho de crianças cursando da 1.^a à 4.^a série do Ensino Fundamental no referido subteste. As autoras demonstraram que o critério de acerto apresentado no manual torna o teste muito permissivo, por considerar como corretas respostas que denotam a construção laboriosa da pronúncia sem o mínimo de automatismo. Além disso, a seleção dos itens parece não ter sido adequada, pois, paralelamente a uma escassez de itens difíceis (o que provocou, na amostra investigada, um efeito de teto na distribuição dos escores), apresentou um excesso de itens fáceis. Este resultado, além de impedir uma discriminação dos leitores mais eficientes daqueles que apresentam problemas de leitura, é capaz de repercutir em prejuízos no processo de avaliação psicológica: como uma grande quantidade de itens deve ser lida a fim de que o indivíduo atinja o nível inferior de classificação no subteste (73% já na 1.^a série), há um desperdício na alocação de recursos de tempo de avaliação ou mesmo na colaboração do avaliando em participar da testagem.

Em relação ao subteste de leitura de palavras isoladas do PROLEC, foram criadas normas de interpretação para crianças cursando da 1.^a à 4.^a série do Ensino Fundamental de escolas particulares, municipais e rurais (de cidades não informadas). O subteste é composto por 30 palavras divididas em termos de estrutura silábica. Não constam, no manual, informações sobre a frequência ou o comprimento das palavras³. Os autores apontam que houve diferenças de série e de tipo de escola no teste, mas estas parecem não se estender ao subteste de leitura (de fato, não fica claro, pelo relato fornecido, quais são os subtestes que discriminaram os grupos de crianças, uma vez que essa informação é dada de uma forma muito geral e os valores estatísticos não foram reportados no manual). De maneira semelhante ao que ocorre com o subteste de leitura TDE, a distribuição dos escores na prova de leitura do PROLEC segue a direção de um efeito de teto (as crianças leem em média 28 palavras já na 1.^a série). Os autores ressaltam a importância do estabelecimento de uma análise qualitativa para complementar as informações obtidas por meio dos dados quantitativos, tarefa que se complica em função da ausência de pareamento das características psicolinguísticas dos itens lexicais (ver nota de rodapé nº 3).

Tendo em vista a importância de se considerar tanto a teoria quanto os parâmetros psicométricos na construção dos testes psicológicos, o presente trabalho, que é parte do projeto de pesquisa de Pinheiro (2007b), tem como objetivo discutir os fundamentos para a criação de uma

¹ Ou seja, pela confirmação das inferências produzidas pela teoria a partir de testes empíricos do desempenho dos avaliandos nos itens dos testes (Gressler, 2004). No caso que estamos lidando, a produção de efeitos na leitura previstos pelo modelo de dupla-rota.

² A. G. S. Capovilla e F. C. Capovilla (1998) estabeleceram normas preliminares para uma tarefa de leitura em voz alta de crianças cursando da pré-escola à 2.^a série do Ensino Fundamental. O trabalho apresenta como problemas a amostra reduzida (76 participantes) e homogênea (foi utilizada apenas uma escola da rede particular). A mesma tarefa passou por uma validação de construto (A. G. S. Capovilla, F. C. Capovilla & Macedo, 1998), mas não foram relatadas análises de itens, de fidedignidade ou de outros índices de validade. Além disso, F. C. Capovilla, Varanda e A. G. S. Capovilla (2006) publicaram a normatização e a validação de uma bateria de testes de competência na leitura de palavras reais e de não-palavras. Tal bateria difere-se da tarefa de reconhecimento de palavras com que estamos lidando aqui por ser constituída de tarefas de leitura silenciosa em que, adicionalmente aos estímulos escritos, são também apresentados estímulos pictóricos. Por essas razões, os subtestes de leitura do TDE e da bateria PROLEC podem ser considerados os únicos testes de leitura atualmente comercializados no Brasil e que apresentam as características psicométricas supracitadas.

³ Fazendo uma análise pela contagem de Pinheiro (1996), observa-se que a maioria das palavras é regular, totalizando 53% dos casos, e uma minoria é governada por regras contextuais (10%). Os itens se dividem de forma homogênea em curtos (menos de 5 letras) e grandes (mais de 5 letras), mas há uma prevalência de itens entre 5 e 7 letras (que totalizam 83%). Por fim, há certo equilíbrio entre palavras de alta e baixa frequência (respectivamente, 40% e 37%), sendo as restantes de frequência média (23%).

tarefa de leitura em voz alta de palavras isoladas destinada a avaliar a habilidade de reconhecimento de palavras de crianças cursando os anos iniciais do Ensino Fundamental. Apresentamos aqui a etapa de análise empírica dos itens que representou o primeiro estágio da seleção dos itens-candidatos a compor a versão final da referida tarefa, bem como ressaltamos o processo que norteou a escolha dos itens que foram aplicados na amostra.

Método

Considerações Éticas

Este trabalho é parte de um projeto maior, intitulado “Avaliação de Competências Psicológicas da População Infanto-Juvenil de Belo Horizonte: Inteligência e Habilidades de Leitura e Escrita” (Pinheiro, 2007b). Este projeto foi aprovado pelo Comitê de Ética em Pesquisa com Seres Humanos (COEP) da Universidade Federal de Minas Gerais ([UFMG] processo n.º ETIC 347/04).

Participantes

A amostra foi composta de 333 crianças do 2º ao 5º ano do Ensino Fundamental⁴ (média de idade = 8,77, $DP = 2,14$; 52% do sexo feminino), oriundas de sete escolas da Rede de Ensino de Belo Horizonte (quatro estaduais, duas municipais e uma particular). O nível socioeconômico médio das famílias das crianças variou de A2 a D (Associação Brasileira de Empresas de Pesquisa [ABEP], 2003). A coleta de dados ocorreu entre agosto de 2006 e março de 2007.

Instrumento

Todos os participantes leram em voz alta uma lista de 323 de palavras de baixa frequência de ocorrência, variando em níveis de regularidade e em número de letras (Pinheiro, 1996, 2007a). Em relação ao nível de regularidade, cada palavra foi classificada de acordo com a correspondência grafema-fonema (direção da leitura) e a correspondência fonema-grafema (direção da escrita). Essa classificação bidirecional de palavras foi introduzida por Pinheiro (2004) e foi testada em dois estudos (Lúcio, Pinheiro, & Nascimento, 2010; Pinheiro, Lúcio, & Silva, 2008). Esses estudos mostram que, no que se refere à precisão, a leitura de escolares sofre um maior efeito da regularidade quando as palavras são classificadas tomando como referência a direção da leitura, enquanto que para o tempo de reação o fator mais importante é a direção da escrita. Assim, como o presente estudo toma como objeto de análise a proporção de acertos nos itens, embora as palavras apresentas-

sem classificação de regularidade bidirecional, as análises foram conduzidas tomando como referência a direção da leitura. Por essa razão, iremos descrever a classificação das palavras apenas desse ponto de vista, que pode ser assim resumida: palavras regulares (R), as quais apresentam relação transparente entre grafema e fonema (p. ex.: cava); palavras regra (RG), que contêm grafemas cujas pronúncias são regidas por regras posicionais, dependentes de contexto ortográfico (p. ex.: caro); palavras irregulares (IR), que contêm grafemas com pronúncias arbitrárias (p. ex.: boba); e palavras irregulares-regra (IRRG), que apresentam simultaneamente um grafema cuja pronúncia é irregular e outro contendo regra contextual (p. ex.: aceso).

No presente estudo, contrariando as recomendações de Parente, Hosogi e Lecours (1997), foram selecionadas apenas palavras de baixa frequência (Pinheiro, 1996). No entanto, tal escolha se justifica por duas razões, uma de ordem empírica e outra de ordem prática. Do ponto de vista empírico, sabemos que o nível de regularidade não afeta a leitura das palavras de alta frequência nem mesmo entre os leitores iniciantes, havendo apenas uma presença restrita deste efeito nas palavras infrequentes (p. ex., C. N. G. Justi & F. R. R. Justi, 2009; Pinheiro, 1995, 2008). Sendo assim, o conhecimento prévio do efeito da frequência na produção de erros e no tempo de reação relatados naquele trabalho norteou a seleção de apenas as palavras de baixa frequência no presente estudo. Do ponto de vista prático, Pinheiro e Rothe-Neves (2001) chamam a atenção para um problema de ordem metodológica na construção de listas pareadas de palavras de baixa e de alta frequência para a avaliação do efeito de regularidade na leitura. Isso ocorre porque o universo de palavras de alta frequência é muito pequeno, tornando inviável o levantamento de um grande número desses itens divididos em termos de regularidade, número de letras e estrutura silábica. Por esses motivos, neste estudo decidiu-se utilizar apenas palavras de baixa frequência e, após a seleção daquelas palavras que se mostrarem mais adequadas em termos de dificuldade e discriminação, será finalmente feito o pareamento com as palavras de alta frequência que constam do corpo de itens original de Pinheiro (1996).

Procedimento de Aplicação

O conjunto de itens foi aplicado por meio de computadores portáteis e dividido em duas partes, a Lista 1 e a Lista 2. Em cada aplicação, havia duas pausas pré-programadas, para o caso de a criança necessitar de descanso. As palavras apareciam no centro da tela do computador em letra *Times New Roman* tamanho 40, escritas na cor branca com fundo preto. As respostas e o tempo de reação das crianças (medido em milésimos de segundo) foram automaticamente gravados pelo computador por meio do programa *Cogwork* (gentilmente cedido por Philip Seymour da Universidade de Dundee, Escócia). Além do conjunto de itens, as crianças foram avaliadas em outras habilidades, mas os resultados não serão relatados aqui.

⁴ De acordo com informações fornecidas pelos pais das crianças, 97% da amostra frequentou a pré-escola. Por esta razão, mesmo tendo a coleta sido realizada em uma época em que não era obrigatória a entrada das crianças na escola aos 6 anos de idade, preferimos adotar a nova nomenclatura para os anos escolares. Desta forma, o 2º ano de escolarização refere-se à antiga 1.ª série, o 3º ano à 2.ª e assim sucessivamente.

Procedimento de Análise dos Dados

Foi conduzida uma análise para investigar a unidimensionalidade da tarefa com base no cálculo da consistência interna (alfa de Cronbach). Para este índice, foram considerados satisfatórios valores maiores do que 0,80. Em seguida, foram investigados os parâmetros de discriminação [por meio do índice D e correlação item-total (I-T)] e de dificuldade (índice p , calculado pela proporção de participantes que acertaram o item). Para o cálculo do índice D , os grupos superior (S) e inferior (I) foram compostos com base no desempenho dos participantes que obtiveram escores totais posicionados aos 27% das extremidades superior e inferior da distribuição, respectivamente. Neste caso, valores acima de 0,30 foram considerados aceitáveis. Por fim, para a dificuldade dos itens, procurou-se seguir o critério de Pasquali (1996), segundo o qual a distribuição deve apresentar a mesma proporção da curva normal. Assim, os níveis de dificuldade foram divididos em cinco faixas [muito difícil ($p < 0,20$), difícil ($0,20 \leq p < 0,40$), médio ($0,40 \leq p < 0,60$), fácil ($0,60 \leq p < 0,80$) e muito fácil ($p \geq 0,80$)], sendo que a maior parte dos itens selecionados

deve apresentar dificuldade média e apenas uma pequena proporção de itens deve ser muito difícil ou muito fácil. As análises estatísticas foram conduzidas no programa estatístico SPSS 16.0 for Windows.

Resultados

Os valores de alfa de Cronbach atingiram níveis altamente satisfatórios, variando de 0,97 na 3ª série a 0,99 na 1ª série e no geral (todas as séries juntas). Os índices de consistência interna obtidos apontam em favor da homogeneidade da tarefa, indicando que o construto avaliado, a habilidade de reconhecimento de palavras isoladas, se mostra unidimensional.

A Tabela 1 apresenta a média dos índices p , D e I-T para cada série e no total. Observe que, mesmo para a correlação item-total, optou-se por utilizar os valores médios ao invés da transformação desses valores em escores z pelo fato de a distribuição ter sido aproximadamente normal para todas as séries e no total (os valores modulares mínimo e máximo para a assimetria e para a curtose foram de 0,40 e 0,65 entre as séries e de 0,15 e 0,53 no total).

Tabela 1

Valores Médios da Dificuldade, Discriminação e Correlação Item-Total nas Listas 1 e 2 e Para Toda a Lista em Cada Série e Para Toda a Amostra

Série	Lista 1			Lista 2			Total		
	p	D	I-T	p	D	I-T	p	D	I-T
1ª	0,73	0,45	0,49	0,72	0,41	0,45	0,72	0,43	0,47
2ª	0,81	0,32	0,45	0,78	0,32	0,42	0,80	0,32	0,43
3ª	0,87	0,22	0,31	0,85	0,23	0,29	0,86	0,23	0,30
4ª	0,88	0,27	0,48	0,86	0,28	0,49	0,87	0,27	0,48
Total	0,82	0,34	0,47	0,81	0,33	0,44	0,81	0,34	0,45

Nota. p = proporção de acertos; D = coeficiente de discriminação; I-T = correlação item-total.

A Tabela 2 mostra a proporção de palavras das Listas 1 e 2 e do total que se distribuem ao longo dos diferentes níveis de dificuldade adotados, bem como dos níveis de discriminação e da consistência interna. Por ela, podemos observar que a maioria das palavras das listas é fácil ($0,60 \leq p < 0,80$), ou muito fácil ($p \geq 0,80$)⁵, e que apenas uma proporção ínfima de itens apresenta um nível extremo de dificuldade ($p < 0,20$). Apesar disso, mais da metade dos itens produziu níveis adequados de discriminação ($D \geq 0,30$), o que indica que o conjunto de itens como um todo foi capaz de separar os indivíduos com competência na leitura de palavras isoladas daqueles indivíduos cuja competên-

cia é limitada. Por fim, observou-se que a maioria dos itens também produziu níveis adequados de correlação item-total ($I-T \geq 0,30$), o que também atesta para a homogeneidade da tarefa.

Devido à pequena quantidade de itens que apresentam índices de dificuldade média a muito alta ($p < 0,60$), todos os itens de ambas as Listas 1 e 2 com estes níveis de dificuldade são reportados na Tabela 3. O índice de discriminação desses itens variou de 0,13 a 0,70. Para as palavras com baixa dificuldade ($0,60 \leq p < 0,80$), foram selecionados os itens com $D \geq 0,50$ e $I-T \geq 0,40$ (valores máximos, respectivamente, de 0,70 e 0,60). Com isso, foram selecionados 34 itens (17 de cada lista). Mantiveram-se, assim, proporções em torno de 45% dos itens originais em ambas as listas para o nível de dificuldade baixo. Das 223 palavras que apresentaram nível de dificuldade muito baixo ($p \geq 0,80$), foram escolhidas aquelas que apresentaram $D \geq 0,35$ e $I-T \geq 0,40$ (com valores máximos, respectivamente, de 0,53 e

⁵ O que refletiu nas altas percentagens de acerto no teste, conforme se observa pela Tabela 1. Em média, considerando todas as séries juntas, a percentagem de acerto no teste foi de 81%. Uma análise completa dos resultados pode ser vista em outro trabalho (ver Lúcio et al., 2010).

0,67). Ao todo, permaneceram 54 itens (24% do total), sendo 30 deles oriundos da Lista 1. Dos 323 itens originais da

tarefa, foram selecionados 112 itens, baseando-se nos critérios acima expostos. A Tabela 3 resume os resultados.

Tabela 2

Distribuição da Proporção de Palavras (%) em Cada Nível de Dificuldade, Discriminação e Correlação Item-Total

Análise estatística	Lista 1	Lista 2	Total
Dificuldade			
$p < 0,20$	0,62	0	0,31
$0,20 \leq p < 0,40$	1,23	1,24	1,24
$0,40 \leq p < 0,60$	3,09	8,70	5,88
$0,60 \leq p < 0,80$	24,07	22,98	23,53
$p \geq 0,80$	70,99	67,08	69,04
Discriminação			
$D < 0,30$	38,27	45,34	41,80
$D \geq 0,30$	61,73	55,66	58,20
Correlação item-total			
I-T $< 0,30$	3,09	8,07	5,57
I-T $\geq 0,30$	96,91	91,93	94,93

Nota. p = proporção de acertos; D = coeficiente de discriminação; I-T = correlação item-total.

Tabela 3

Itens Selecionados por Meio da Análise, Separados em Função do Nível de Dificuldade (p) e da Lista de Aplicação

Dificuldade	R-1	R-2
$p < 0,20$	ampola	—
$0,20 \leq p < 0,40$	aceso, exceto	tola, boxe
$0,40 \leq p < 0,60$	serelepe, sarjeta, foco, cera, humilde	palhoça, força, catorze, roxa, saxofone, zebu, despesa, monarca, novelo, guizo, bodoque, rasa, maroto, careta
$0,60 \leq p < 0,80$	excesso, exposto, gengiva, freguesa, bondosa, caroço, atleta, colo, manhoso, bezerro, caçula, cigana, celeste, marreco, nublado, vasilha, cadela	externo, colete, rabanete, cometa, cevada, cesto, enxoval, fofo, delicada, serrote, deserto, doutora, luminosa, reflexo, mariposa, defesa, bexiga
$p \geq 0,80$	cereja, cartela, espiral, veste, covarde, esquerdo, farda, recurso, teimoso, repleto, diretora, aflito, faxina, morena, dengoso, capela, feminina, carroça, negra, pepino, vaidosa, peteca, gemada, coberta, uniforme, combate, fiapo, terço, sapeca, selva.	cantora, nervosa, caxumba, retalho, jeitoso, cabana, universo, famoso, moderna, favela, preguiça, adulta, brasa, frota, miolo, cartola, trevo, honesto, tombo, bandeja, tapete, espanto, horta, filtro.

Nota. R-1 (palavras obtidas pela análise de itens da Lista 1); R-2 (palavras obtidas pela análise de itens da Lista 2).

A Tabela 4 mostra a configuração final da distribuição dos itens selecionados em termos das quatro classificações de regularidade para a leitura e do nível de dificuldade, para as formas reduzidas das Listas 1 e 2 e para o conjunto total de itens. Por esta tabela, pode-se observar que em ambas as formas os itens de dificuldade alta e muito alta

se concentram entre as palavras irregulares para a leitura (IR e IRRG) e que essas palavras também são predominantes entre os níveis de dificuldade média. As palavras regulares (R e RG) só aparecem entre os itens dificuldade média, mas mesmo assim não são prevalentes em nenhum dos níveis de dificuldade aqui especificados.

Tabela 4

Porcentagens de Palavras em Cada Classificação de Regularidade para os Níveis de Dificuldade (p) Especificados

Dificuldade/Lista	R	RG	IR	IRRG
R-1				
$p < 0,20$	-	-	100	-
$0,20 \leq p < 0,40$	-	-	50	50
$0,40 \leq p < 0,60$	20	-	40	40
$0,60 \leq p < 0,80$	12	18	41	29
$p \geq 0,80$	37	7	43	13
R-2				
$p < 0,20$	-	-	-	-
$0,20 \leq p < 0,40$	-	-	100	-
$0,40 \leq p < 0,60$	14	14	50	21
$0,60 \leq p < 0,80$	6	12	35	47
$p \geq 0,80$	33	4	46	17
Total				
$p < 0,20$	-	-	100	-
$0,20 \leq p < 0,40$	-	-	75	25
$0,40 \leq p < 0,60$	16	11	47	26
$0,60 \leq p < 0,80$	9	15	44	32
$p \geq 0,80$	35	6	44	15

Nota. R-1 e R-2 (palavras obtidas, respectivamente, pela análise de itens da Lista 1 e da Lista 2); palavras regulares (R), regra (R) irregulares (IR) e irregulares-regra (IRRG).

Discussão

O presente estudo relatou as etapas intermediárias conduzidas para a investigação psicométrica de uma tarefa de leitura em voz alta de palavras isoladas, a qual será destinada à avaliação de alunos cursando do 2º ao 5º ano do Ensino Fundamental da cidade de Belo Horizonte. Especificamente, discutimos os fundamentos teóricos da escolha da lista de palavras utilizada e reportamos uma análise de itens a partir da qual foram selecionados os itens-candidatos a fazer parte da referida tarefa.

A análise de itens demonstrou que, do ponto de vista de acurácia, o conjunto de itens testado não apresentou muitas dificuldades para os leitores, sendo que a grande maioria dos itens mostrou-se fácil ou muito fácil para toda a amostra. Isso corrobora com questões levantadas por pesquisadores a respeito do caráter regular ou previsível de nossa língua (p. ex., Parente, Silveira, & Lecours, 1997), que impõe pouca ou nenhuma dificuldade mesmo para os leitores iniciantes. Este fato torna-se evidente quando observamos que apenas cinco itens apresentaram índice de acerto inferior a 0,40 (ver Tabela 3). Considerando que mais de 50% das palavras foram classificadas como irregulares para a leitura, a porcentagem de itens que impuseram dificuldades para a amostra nos faz refletir sobre a classificação de regularidade utilizada no português brasileiro. Diante desse fato, pelo menos no que se refere à precisão na leitura, talvez fosse mais apropriado se falar em ambiguidade ou em inconsistência do que propriamente

em “irregularidade” entre as palavras do português que apresentam pronúncias que não são regulares ou não são previsíveis por regras contextuais. De fato, Silva e Rothe-Neves (2002) demonstraram que o grafema <x> apresenta uma pronúncia consistente (lido como /f/) em 75% das vezes, ou seja, a criança que desconhece qualquer outra pronúncia do <x> tem 75% de chance de acertá-la na leitura de uma palavra contendo esse grafema⁶. Isso não caracteriza uma irregularidade, tal como se observa em outros idiomas, como o inglês.

A despeito desse fato, a onipresença de itens irregulares entre as palavras mais difíceis (ver na Tabela 4 palavras com índice $p < 0,40$) e a quantidade marcante de itens com esse tipo de regularidade (69%) que apresentou os níveis satisfatórios de discriminação e de dificuldade atenta para o poder do efeito de regularidade no português brasileiro. A presença deste efeito na seleção dos itens demonstra que a dificuldade que a palavra impõe ao leitor iniciante está intimamente relacionada ao processo de leitura que é utilizado por ele no momento da realização da tarefa. Por este motivo, tal fato corrobora de forma indireta com a aplicabilidade do modelo de dupla-rota no nosso idioma, uma vez que os itens irregulares, quando lidos pela rota fonológica, podem estar sujeitos a erros de pro-

⁶ Pelo menos entre as palavras de alta frequência, de onde foi retirada proporção feita por Silva e Rothe-Neves (2002) da representação de fonemas por grafemas da nossa língua.

núncia, o que justamente caracteriza a dificuldade dos itens com os quais estamos lidando.

Apesar de ser plausível a hipótese de que o efeito de regularidade está relacionado ao nível de dificuldade da palavra, outra possibilidade deve ser levada em consideração. É possível que outras características das palavras além da regularidade (como o número de palavras vizinhas ou a presença de um grafema muito incomum), que não foram estudadas, possam ter interferido nos resultados apresentados. Isso fica mais evidente a partir dos índices de correlação item-total que as palavras de dificuldade extremamente alta apresentaram, os quais ficaram abaixo de 0,30 na maioria dos casos (60%). Este fato demonstra que habilidades diferentes do reconhecimento de palavras⁷ possam ter interferido no sucesso na leitura desses itens, uma vez que os índices de consistência interna que aqui foram obtidos (dados pelo alfa de Cronbach) apontaram para uma forte unidimensionalidade do construto investigado.

As questões relativas à dificuldade das palavras no português, aliadas à configuração final dos itens-candidatos obtidos pela análise, nos fazem refletir sobre a aplicação em nosso idioma das recomendações de Pinheiro e Rothe-Neves (2001) em relação às características das listas de palavras utilizadas para uma avaliação cognitiva da leitura. Conforme dito na introdução, para os autores, com base na teoria de reconhecimento de palavras, é essencial que os itens de uma lista construída para fins de permitir uma avaliação dos processos de leitura sejam pareados em termos de regularidade, frequência e número de letras a fim de se extrair conclusões a respeito do processamento utilizado pela criança na leitura. No entanto, como demonstramos, em uma lista de centenas de itens de baixa frequência de ocorrência, apenas 1,5% apresentou índices de dificuldade mais elevada para distinguir as habilidades de leitura de crianças em fase de alfabetização. Isso mostra que as recomendações dos autores parecem se aplicar mais aos estudos que visam identificar as estratégias utilizadas pelos leitores durante a leitura do que a um trabalho em que a metodologia se liga aos procedimentos psicométricos de construção de um teste, cujo objetivo é discriminar leitores competentes dos não-competentes. Os resultados aqui apresentados mostram que o pareamento de palavras que se diferem em termos de frequência⁸ e

regularidade pode não ter relevância do ponto de vista normativo, sendo talvez importantes apenas os índices psicométricos de dificuldade e de discriminação. Obviamente, o referido pareamento permanece relevante no momento de se investigar a validação do construto da tarefa e também na condução de uma avaliação dos processos de leitura que mudam de acordo com a competência do leitor e de seu estágio de desenvolvimento nessa habilidade. No entanto, a inclusão de palavras de alta frequência de ocorrência entre os itens-candidatos que foram aqui obtidos pode apenas acrescentar informações redundantes e que não contribuem para a avaliação das diferenças individuais na habilidade geral de leitura requerida pelo teste⁹.

A questão aqui levantada acerca da aplicabilidade do pareamento de itens em termos de regularidade e frequência, para a avaliação normativa da habilidade de reconhecimento de palavras, pode ser esclarecida pela diferenciação feita por Embretson (1983) a respeito das pesquisas de validação de construto, referidas pela autora como pesquisas de representação do construto e pesquisas de intervalo nomotético. No primeiro caso, o objetivo é identificar os mecanismos teóricos que estão envolvidos no desempenho de uma tarefa, ou seja, há uma busca pela decomposição da tarefa em etapas irreduzíveis. Desta forma, no paradigma do processamento de informação, a representação do construto se relaciona à dependência que o desempenho nas tarefas possui em relação aos processos, estratégias e conhecimento anteriores do indivíduo. Apesar de esta dependência poder ser um aspecto essencial para o desempenho na tarefa, isso não implica que as pessoas irão necessariamente variar em termos deste desempenho. Em outras palavras, a pesquisa de representação do construto está interessada com a variabilidade das tarefas de um teste e não necessariamente com as diferenças individuais produzidas pelas mesmas. No caso da presente pesquisa, por exemplo, em um conjunto de itens com centenas de palavras consideradas mais difíceis, por serem pouco familiares, ocorreu uma proporção geral de 81% de acertos em uma amostra composta por leitores competentes e menos habilidosos. Desta forma, por mais que supostamente os processos de leitura utilizados por ambos os tipos de leitores sejam diferentes (ou seja, o uso de processamento lexical no primeiro grupo e um apego à estratégia fonológica no segundo); esta diferença não refletiu em uma grande variabilidade nos escores dos participantes, o que é representado pela assimetria na distribuição dos escores do banco de itens.

⁷ O conhecimento da pronúncia de grafemas infrequentes pode ser um desses fatores. O grafema <xc>, por exemplo, apesar de apresentar a pronúncia regular /s/, por ser extremamente infrequente (Silva & Rothe-Neves, 2002), pode não ser lido corretamente por leitores iniciantes ou pouco habilidosos.

⁸ Isso pode ser afirmado por inferência, pois, como dito no início desse parágrafo, apenas uma minoria dos itens presentes na lista original apresentou alguma dificuldade para a amostra. Considerando que as palavras de baixa frequência são mais difíceis de serem lidas do que as palavras familiares, uma vez que sua representação no léxico mental é mais fraca (Coltheart et al., 2001), é esperado que as palavras de alta frequência imponham uma dificuldade menor ainda para os leitores.

⁹ É importante ressaltar que as hipóteses aqui levantadas se baseiam em uma análise que considerou apenas a precisão ou proporção de acertos na leitura. É possível que resultados diferentes sejam obtidos se a seleção dos itens também se basear no tempo de leitura dos participantes e se forem analisados via TRI que, em comparação com a TCT, oferece uma melhor caracterização e seleção de itens.

A pesquisa de intervalo nomotético, por outro lado, se refere à rede de inter-relações que se pode estabelecer entre um teste e outras medidas (que meçam o mesmo construto ou mesmo construtos diferentes). Nesse sentido, este tipo de pesquisa se preocupa com diferenças individuais e se interessa pela variabilidade dos escores dos participantes nas tarefas.

Pode-se dizer que os estudos experimentais que objetivam observar os efeitos diversos na leitura de palavras (e, portanto, validar os modelos que distinguem os processos e as etapas utilizados durante o ato de ler) constituem pesquisas de representação do construto. Enquanto isso, as pesquisas voltadas para a construção de testes e o estabelecimento de normas que são utilizadas para classificar os indivíduos em diferentes níveis de habilidade são caracterizadas como pesquisas de intervalo nomotético. Desta forma, ambos os tipos de pesquisa (de representação do construto e de intervalo nomotético) irão buscar fontes distintas de validação que são independentes umas das outras. Aplicando essa distinção para as questões levantadas acima, pode-se dizer que o pareamento das palavras em termos de características psicolinguísticas é relevante para a validação da pesquisa experimental na área da leitura de palavras isoladas, enquanto que a sua separação em termos de dificuldade e discriminação é importante para as comparações intragrupo que são feitas na mesma área.

Em síntese, a partir da análise reportada, observou-se que a seleção de itens em termos de níveis de dificuldade e de discriminação gerou uma proporção desigual de palavras regulares e irregulares entre os itens-candidatos¹⁰. Se este padrão se mantiver na versão final do teste, este poderia ser útil apenas para uma análise das diferenças entre os testandos em termos do que é esperado em seu desempenho em função do ano de escolarização. Isso implica que esta versão não permitiria a extração de inferências a respeito dos processamentos utilizados na leitura pelos participantes, conforme seria esperado se as palavras fossem pareadas em termos de regularidade, extensão e frequência. Neste caso, o alcance de um objetivo não seria obtido sem o sacrifício do outro. Dito em outros termos, a criação de um teste de palavras isoladas capaz de apresentar variabilidade suficiente a ponto de distinguir as habilidades de indivíduos em níveis diferentes de desenvolvimento poderá exigir a desconsideração do pareamento de palavras que é previsto pela teoria. De maneira análoga, a criação de um instrumento capaz de testar os componentes do modelo de dupla-rota poderia ser alcançada à custa da redução da variabilidade das respostas dos participantes (representada pela entrada no teste de itens mais fáceis – regulares e curtos – passíveis de serem lidos sem dificuldade pelos leitores menos competentes).

Para vencer essas dificuldades, uma nova análise foi conduzida pelo mesmo grupo de pesquisa introduzindo-se a variável tempo de reação para o cálculo dos índices de dificuldade e discriminação, via teoria de resposta ao item. Em função do uso do tempo de reação para a seleção dos itens, e como os estudos de Lúcio et al. (2010) e de Pinheiro et al. (2008) demonstraram que a direção da classificação de regularidade interfere na produção de erros de no tempo de reação, Pinheiro (2011) optou por escolher as palavras que apresentam classificação idêntica tanto para a leitura quanto para a escrita, a fim de se controlar o efeito que esta variável apresenta nas variáveis dependentes do estudo (por exemplo, uma palavra foi classificada como regular considerando-se tanto o ponto de vista da leitura quanto da escrita, como *cava*, eliminando-se da lista as palavras regulares para a leitura, mas irregulares para a escrita, como *caça*. Igualmente, para uma palavra ser considerada irregular ela deve ser irregular nas duas direções, como é o caso de *relva*, e assim por diante). Acreditamos que esta medida poderá fortalecer o efeito de regularidade nos itens e contribuir para um aumento da dificuldade apresentada pelas palavras irregulares. Os itens obtidos por meio deste procedimento estão sendo atualmente testados em uma nova amostra de participantes. Se demonstrado eficaz, este procedimento poderá incrementar a variabilidade da distribuição dos escores e, assim, aumentar a oferta de itens regulares e curtos entre aqueles com maior nível de dificuldade.

Conclusão

A título do que ocorreu em trabalhos anteriores (Cuetos et al., 1998; Lúcio et al., 2009), o conjunto de itens analisado não apresentou dificuldades para a amostra investigada. Estes resultados nos fazem refletir sobre a tradicional classificação de regularidade das palavras no português e sobre a construção de testes para a avaliação da leitura de palavras isoladas entre leitores iniciantes, a qual deve considerar a peculiaridade de nosso idioma. Particularmente, acreditamos ser necessária a construção de diferentes tipos de instrumentos voltados para uma avaliação completa da leitura. Assim, parece ser plausível a criação de instrumentos que visem à produção de uma maior variabilidade intragrupo (calculada nas características psicométricas de dificuldade e de discriminação dos itens), assim como de instrumentos que busquem avaliar os componentes do modelo de dupla-rota (em que se considera o pareamento das palavras em termos de características psicolinguísticas). Os testes atualmente comercializados no Brasil parecem não atender a nenhum desses objetivos. O uso do tempo de reação como uma variável dependente na construção dos itens, para além da precisão, combinado com uma classificação de regularidade de palavras que imponha um nível maior de dificuldade para a leitura, se abrem como uma nova possibilidade dentro desta configuração de resultados.

¹⁰ Esta desproporção também se estendeu para itens grandes e curtos, uma vez que apenas 25 palavras (20,5% da amostra total de itens-candidatos) apresentaram 4 ou 5 letras.

Referências

- Amendola, M. F. (2006). *Panorama da história dos testes psicológicos no Brasil e críticas atuais*. Retrieved August 05, 2010, from <http://www.canalpsi.psc.br>
- Associação Brasileira de Empresas de Pesquisa. (2003). *Crítério de classificação econômica Brasil*. Retrieved August 15, 2010, from <http://www.abep.org>
- Capovilla, A. G. S., & Capovilla, F. C. (1998). Desenvolvimento de leitura e escrita de pré 3 à 2ª série: Tabelas preliminares de normatização de uma lista de itens psicolinguísticos. *Ciência Cognitiva: Teoria, Pesquisa e Aplicação*, 2(4), 821-840.
- Capovilla, A. G. S., Capovilla, F. C., & Macedo, E. C. (1998). Validação do software CronosFonos para a análise de tempo de reação, duração e frequência de segmentação locucionais na leitura em voz alta de itens isolados. *Ciência Cognitiva: Teoria, Pesquisa e Aplicação*, 2(3), 253-340.
- Capovilla, F. C., Varanda, C., & Capovilla, A. G. S. (2006). Teste de Competência de Leitura de Palavras e Pseudopalavras: Normatização e validação. *Revista de Psicologia da Vetor Editora*, 7(2), 47-59.
- Chadha, N. K. (2009). *Applied psychometry*. Delhi, India: Sage.
- 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.
- Conselho Federal de Psicologia. (2003). *Resolução 02/2003*. Retrieved July 05, 2010, from <http://www.pol.org.br>
- Cuetos, F., Rodrigues, B., & Ruano, E. (1998). *Provas de avaliação dos processos de leitura – PROLEC: Manual* (S. A. Capellini, A. M. Oliveira, & F. Cuetos, adaptação brasileira). São Paulo, SP: Casa do Psicólogo.
- Embretson, S. (1983). Construct validity: Construct representation versus nomothetic Span. *Psychological Bulletin*, 92(1), 179-197.
- Gressler, L. A. (2004). *Introdução à pesquisa: Projetos e relatórios* (2. ed.). São Paulo, SP: Loyola.
- Justi, C. N. G., & Justi, F. R. R. (2009). Os efeitos de lexicalidade, frequência e regularidade na leitura de crianças falantes do português brasileiro. *Psicologia: Reflexão e Crítica*, 22(2), 163-172.
- Levy, J., Pernet, C., Treserras, S., Boulanouar, K., Aubry, F., Démonet, J., et al. (2009). Testing for the Dual-Route Cascade Reading Model in the Brain: An fMRI effective connectivity account of an efficient reading style. *Plos One*, 4(8), 1-13.
- Lúcio, P. S., Pinheiro, A. M. V., & Nascimento, E. (2009). O impacto da mudança no critério de acerto na distribuição dos escores do subteste de leitura do Teste de Desempenho Escolar. *Psicologia em Estudo*, 14(3), 593-601.
- Lúcio, P. S., Pinheiro, A. M. V., & Nascimento, E. (2010). A influência de fatores sociais, individuais e linguísticos no desempenho de crianças na leitura em voz alta de palavras isoladas. *Psicologia: Reflexão e Crítica*, 23(3), 496-505.
- Noronha, A. P. P., Primi, R., & Alchieri, J. C. (2005). Instrumentos de avaliação mais conhecidos/utilizados por psicólogos e estudantes de psicologia. *Psicologia: Reflexão e Crítica*, 18(3), 390-401.
- Parente, M. A. M. P., Hosogi, M. L., & Lecours, A. R. (1997). Conduta clínica. In A. R. Lecours & M. A. M. Parente (Eds.), *Dislexia: Implicações do sistema de escrita do português* (pp. 85-105). Porto Alegre, RS: Artes Médicas.
- Parente, M. A. M. P., Silveira, A., & Lecours, A. R. (1997). As palavras do português escrito. In A. R. Lecours & M. A. M. Parente (Eds.), *Dislexia: Implicações do sistema de escrita do português* (pp. 41-55). Porto Alegre, RS: Artes Médicas.
- Pasquali, L. (1996). *Teorias e métodos de medida em ciências do comportamento*. Brasília, DF: Laboratório de Medida em Ciências do Comportamento, Universidade de Brasília.
- Pinheiro, A. M. V. (1995). Reading and spelling development in Brazilian Portuguese. *Reading and Writing: An Interdisciplinary Journal*, 7, 111-138.
- Pinheiro, A. M. V. (1996). *Contagem de frequência de ocorrência de palavras expostas a crianças de 1ª à 4ª série do Ensino Fundamental*. São Paulo, SP: Associação Brasileira de Dislexia.
- Pinheiro, A. M. V. (2004). *Banco de palavras de alta e baixa frequência de ocorrência, para crianças brasileiras da 1ª à 4ª série do Ensino Fundamental, classificadas em termos de estrutura silábica, número de letras e regularidade para leitura e para escrita* (Relatório FAPEMIG No. DC/SOT 1202/2004). Belo Horizonte, MG: Universidade Federal de Minas Gerais, Departamento de Psicologia.
- Pinheiro, A. M. V. (2007a). Anexo 2. In I. Sim-Sim & F. L., Vianna, *Para a avaliação do desempenho de leitura* (pp. 121-131). Lisboa, Portugal: Gabinete de Estatística e Planeamento da Educação. Retrieved September 10, 2010, from <http://www.gepe.min-edu.pt>
- Pinheiro, A. M. V. (2007b). *Avaliação de competências psicológicas da população infanto-juvenil de Belo Horizonte: inteligência e habilidades de leitura e escrita* (Relatório nº SHA-APQ-01914-09). Belo Horizonte, Universidade Federal de Minas Gerais.
- Pinheiro, A. M. V. (2008). *Leitura e escrita: Uma abordagem cognitiva* (2. ed.). Campinas, SP: Livro Pleno.
- 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, S. Capellini, & R. Mousinho, *Dislexia: Novos temas, novas perspectivas*. Rio de Janeiro, RJ: Editora Wak.
- Pinheiro, A. M. V., Lúcio, P. S., & Silva, D. M. R. (2008). O efeito de regularidade grafema-fonema e fonema-grafema na leitura em voz alta de palavras isoladas. *Psicologia: Teoria e Prática*, 10(2), 16-20.
- Pinheiro, A. M. V., & Rothe-Neves. (2001). Avaliação cognitiva da leitura: As tarefas de leitura em voz alta e ditado. *Psicologia: Reflexão e Crítica*, 14(2), 399-408.
- Silva, D. M. R., & Rothe-Neves, R. (2002). O conceito de foneticidade e a relação grafema-fonema no português brasileiro [Resumo]. In *Anais da Semana de Iniciação Científica da UFMG* (p. 10). Belo Horizonte, MG: Universidade Federal de Minas Gerais.
- Stein, L. M. (1994). *TDE: Teste de Desempenho Escolar: Manual para aplicação e interpretação*. São Paulo, SP: Casa do Psicólogo.
- Urbina, S. (2007). *Fundamentos da testagem psicológica* (C. Dornelles, Trad.). Porto Alegre, RS: Artmed. (Original work published 2004).

Recebido: 04/11/2010
1ª revisão: 27/09/2011
Aceite final: 30/09/2011