

Temas em Psicologia

ISSN: 1413-389X

comissaoeditorial@sbponline.org.br

Sociedade Brasileira de Psicologia

Brasil

Silva Lúcio, Patrícia; Vieira Pinheiro, Ângela Maria
Novos Estudos Psicométricos para o Subteste de Leitura do Teste de Desempenho
Escolar
Temas em Psicologia, vol. 22, núm. 1, abril, 2014, pp. 109-119
Sociedade Brasileira de Psicologia
Ribeirão Preto, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=513751527009>

- ▶ Como citar este artigo
- ▶ Número completo
- ▶ Mais artigos
- ▶ Home da revista no Redalyc

Novos Estudos Psicométricos para o Subteste de Leitura do Teste de Desempenho Escolar

Patrícia Silva Lúcio¹

*Departamento de Psicologia e Psicanálise da Universidade Estadual de Londrina,
Londrina, Paraná, Brasil*

Ângela Maria Vieira Pinheiro

*Departamento de Psicologia da Universidade Federal de Minas Gerais,
Belo Horizonte, Minas Gerais, Brasil*

Resumo

O estudo relata a análise dos itens do subteste de leitura do Teste de Desempenho Escolar (Stein, 1994), aplicado em estudantes do 2º ao 5º ano de escolarização da cidade de Belo Horizonte ($N = 341$). Tomando dois critérios de correção – um apresentado no manual (A1) e outro que considerou incorretas respostas de silabação e de correção espontânea (A2) – foi selecionado um conjunto de itens para cada um destes. Apenas 28,6% dos itens originais apresentaram índices adequados de discriminação pelo Critério A1 e 48,6% pelo A2. A proporção de acertos para ambos os critérios foi elevada. Normas intragrupo aos novos conjuntos de itens foram fornecidas. Os resultados indicam que o Critério A2 produziu resultados mais promissores do que o A1.

Palavras-chave: Teste de Desempenho Escolar, análise de itens, leitura, critério.

New Psychometric Studies for the Reading Subtest of the Test of School Performance

Abstract

This study reports the item analysis of the reading subscale of the School Performance Test (Stein, 1994) answered by 341 children from 2nd to 5th grades of Belo Horizonte city. Two sets of items were selected taking as reference different criteria of correction – one presented in the test manual (A1) and another, which considered syllabication and self-correction as incorrect responses. Only 28.6% of the original items presented suitable discrimination indices by the A1 Criterion and 48.6% by the A2. The accuracy for both criteria was high. For each new set of items, intra-group norms were offered. The data indicate that the A2 Criterion produced more promising results than that of A1.

Keywords: School Performance Test, item analysis, reading, criterion.

¹ Endereço para correspondência: Departamento de Psicologia e Psicanálise, Universidade Estadual de Londrina, Rodovia Celso Garcia Cid, Pr 445 Km 380, Campus Universitário, Cx. Postal 10.011, Londrina, PR, Brasil 86057-970. E-mail: pslucio@gmail.com

Agradecimentos: Este trabalho teve apoio da Fundação de Amparo à Pesquisa do estado de Minas Gerais (Fapemig) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Nuevos Estudios Psicométricos para el Subtest de Lectura de la Prueba de Rendimiento Escolar

Resumen

El estudio reporta la análisis de los ítems de la subprueba de lectura de la Prueba de Rendimiento Escolar (Stein, 1994), que se aplicó a estudiantes de 2º a 5º año de la Red de Enseñanza de Belo Horizonte ($N = 341$). Tomando dos criterios de corrección (uno que se presenta en el manual (A1) y otro (A2) que consideró las respuestas de silabación y de corrección espontánea como incorrectas) se ha seleccionado un conjunto de ítems para cada criterio. Sólo el 28,6% de los ítems originales han presentado niveles adecuados de discriminación por el criterio A1, y 48,6% por el A2. La proporción de respuestas correctas para ambos criterios se mostró alta. Se proporcionaron normas intragrupos para los nuevos conjuntos de ítems. Los resultados indican que el Criterio A2 produce resultados más prometedores que A1.

Palabras clave: Prueba de Rendimiento Escolar, análisis de ítems, lectura, criterio.

As normas constituem uma das maneiras de dar significado aos escores dos testes, sendo apropriadas quando o objetivo é criar um grupo de referência para comparar os indivíduos nas habilidades ou traços por eles avaliados (Urbina, 2007). Para serem válidas, as normas necessitam de constante revisão e ampliação, uma vez que efeitos geracionais podem interferir na interpretação dos resultados. É o caso, por exemplo, do *Efeito Flynn*, que se refere ao aumento dos escores dos sujeitos nos testes de inteligência ao longo do tempo, fazendo com que os dados normativos produzidos em uma determinada época possam não servir para as gerações seguintes (Flynn, 2006). Além disso, o aperfeiçoamento e a revisão periódica dos instrumentos e técnicas de avaliação psicológica constituem exigências do Conselho Federal de Psicologia (CFP, 2003), o que ressalta a importância de iniciativas voltadas para esse fim.

O grau em que as normas de fato correspondem ao significado pretendido pelas interpretações dos escores dos testes depende intrinsecamente da qualidade dos seus itens. Segundo Urbina (2007), a análise dos itens constitui um conjunto de procedimentos utilizados na construção e o desenvolvimento dos testes para avaliar a qualidade e as características dos itens que irão compô-lo. Ela envolve tanto procedimentos quantitativos quanto qualitativos, sendo que os primeiros se referem à análise psicométrica, enquanto que os segundos envolvem julgamentos a

respeito da adequação dos itens ao construto que está sendo investigado.

Este trabalho trata da investigação da qualidade dos itens e da revisão e ampliação de normas de comparação intragrupos de uma prova muito conhecida no contexto brasileiro. Trata-se do subteste de leitura do Teste de Desempenho Escolar – TDE (Stein, 1994), que foi criado há quase duas décadas com o objetivo de avaliar a capacidade básica de decodificação. Ele é composto por 70 palavras isoladas que variam em termos de frequência, regularidade e extensão, mas não há um emparelhamento destas características (voltaremos a esse ponto adiante no texto). A tarefa faz parte do primeiro instrumento brasileiro de avaliação das habilidades acadêmicas a apresentar normas para nossa população, além de estudos empíricos atestando para a validade e fidedignidade dos escores. Talvez por esta razão o TDE seja um instrumento amplamente utilizado em pesquisas que buscam avaliar o desempenho acadêmico de crianças cursando os anos iniciais de escolarização (Lúcio, 2008).

Além das implicações discutidas anteriormente, outra motivação para a condução do presente estudo é o fato de haver razões para considerarmos que a qualidade e distribuição dos itens do subteste de leitura do TDE precisa ser revista. Uma destas razões se deve aos resultados apresentados pelo estudo de Lúcio, Pinheiro e Nascimento (2009) que aplicaram a referida tarefa em escolares do 2º ao 5º ano do Ensino Fundamental

da cidade de Belo Horizonte. Com o intuito de induzir uma maior variabilidade na distribuição dos escores e de evitar o efeito de teto, normalmente encontrado na medida sob consideração (ver Lúcio, 2008 para uma revisão), as autoras conferiram aos seus itens dois tipos de escores: um, conforme os critérios de classificação de erros apresentados no manual, e outro, em que a esses critérios foram acrescentados dois tipos de erros: respostas de silabação explícita (leitura de palavras com pausas entre as sílabas, p. ex., a palavra *cavalo* lida como “ca-va-lo”) e de correção espontânea (leitura incorreta, seguida de sua correção). Com essa mudança de critério de acerto, as autoras obtiveram uma maior variabilidade na distribuição dos escores, mas o efeito de teto permaneceu, embora de forma mais atenuada do que na versão original. Em síntese, o estudo de Lúcio et al. (2009) mostrou que, sobretudo em sua versão original, o subteste de leitura do TDE é pouco discriminativo e apresenta problemas de distribuição de itens, já que é composto por uma grande quantidade de itens fáceis e uma proporção pequena de itens mais difíceis, dificultando a discriminação das habilidades dos leitores mais competentes e dos mais experientes.

Paralelamente a esta questão psicométrica, a escolha dos itens também parece não ter sido apropriada do ponto de vista teórico. Assim, uma análise qualitativa mostrou que o subteste não é adequado para uma avaliação cognitiva da leitura de acordo com os modelos correntes (p. ex., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Seidenberg, 2006), uma vez que há uma discrepância da quantidade de itens em função de suas características psicolinguísticas. Ou seja, há uma predominância de itens baixa frequência, regulares e longos, o que impede a verificação, respectivamente, dos efeitos de frequência, de regularidade e de extensão². Em síntese, o sub-

teste parece ter dificuldades em cumprir tanto o critério psicométrico (pelo baixo nível de dificuldade dos itens), quanto o teórico (pela impossibilidade de verificação dos processos que são utilizados pelo leitor para a realização da leitura, em função da ausência de emparelhamento das características psicolinguísticas dos itens).

Desse modo, o presente trabalho tem como principal objetivo conduzir uma análise dos itens do subteste de leitura do TDE, de modo a adequá-lo aos parâmetros psicométricos para a construção de testes (*American Educational Research Association [AERA], American Psychiatry Association [APA], & National Council on Measurement in Education [NMCE]*, 1999). A análise busca selecionar os itens do subteste que apresentam índices adequados de dificuldade e de discriminação, conforme os paradigmas estatísticos propostos pela Teoria Clássica dos Testes. A partir dos dados aqui obtidos, novas normas serão oferecidas para crianças cursando do 2º ao 5º ano de escolarização para o contexto da cidade de Belo Horizonte/MG.

A título do que foi realizado no estudo de Lúcio et al. (2009), serão conduzidas as análises dos itens e reportadas as normas considerando-se dois critérios de acerto, o apresentado no manual do teste (Critério A1) e o proposto por Lúcio et al. (Critério A2). Esta divisão tem dois objetivos distintos: (a) verificar se a seleção dos itens irá diferir segundo um critério ou outro, da mesma forma que as autoras demonstraram que a introdução de alterações no critério de correção produziu mudanças na distribuição dos escores do subteste e (b) preservar as diretrizes propostas por Stein (1994) na versão original do teste, às quais os usuários estão familiarizados e que ainda representa a forma corrente de atribuição dos escores.

Método

Amostra

Participaram do estudo 341 crianças (53% do sexo feminino), cursando do 2º ao 5º ano do

² De acordo com a contagem de Pinheiro (1996), e utilizando como referência o grupo de crianças do 2º ano de escolarização, a maioria dos itens do subteste (79%) apresenta baixa frequência de ocorrência. Quando se utiliza a contagem para o 4º ano, a proporção sobe para 83%. Além disso, o subteste é composto majoritariamente por itens grandes (65,7% com mais de seis letras) e regulares (70%,

dos quais 44% apresentam correspondência gráfe-ma-fonema independente de contexto).

Ensino Fundamental de seis escolas públicas ($N = 280$) e uma escola particular ($N = 61$) da cidade de Belo Horizonte. A média de idade foi de nove anos (mínimo de seis e máximo de 12 anos) e o nível socioeconômico variou de A2 a D (Associação Brasileira de Empresas de Pesquisa [ABEP], 2003).

Instrumentos

Além do subteste de leitura do TDE, as crianças responderam às Matrizes Progressivas de Raven (Raven, Raven, & Court, 1993) e à escala verbal do WISC-III (Wechsler, 1991) e leram uma lista de palavras isoladas (Pinheiro, 1996/2007), assuntos que não serão tratados aqui.

Procedimentos de Coleta de Dados

A presente pesquisa foi aprovada pelo Conselho de Ética em Pesquisa com Seres Humanos da UFMG (processo nº ETIC 347/04). Participaram do estudo apenas as escolas e alunos cujos responsáveis assinaram o termo de consentimento livre e esclarecido. As crianças foram testadas individualmente em ambiente fornecido pela escola e em horário de aula.

Procedimentos de Análise dos Dados

Para a análise dos itens, serão reportados os índices de dificuldade (p ou proporção de acerto no teste), discriminação (D , ou diferença entre os escores dos sujeitos que compõem os 27% mais habilidosos e os 27% menos habilidosos), a correlação item-total e o alfa de Cronbach. Para esta última medida, serão considerados adequados valores acima de 0,80. Em função do tamanho reduzido da amostra de crianças da escola particular, todas as análises serão conduzidas com a amostra total.

Resultados

Análise Exploratória

Pelo Critério A1, houve uma média de 87,5% de acerto nos itens do subteste ($DP = 12,6\%$), enquanto que pelo Critério A2 a média de acertos foi de 75,6% ($DP = 21,4\%$). Apesar

nas quatro itens pelo Critério A1 apresentaram uma proporção de acertos menor do que 60% (hipócrita, vangloriar, excepcional, saguões), enquanto que 13 itens atingiram essa proporção de acertos pelo Critério A2 (hospedaria, rescindido, aeronáutica, repugnante, hipócrita, perseverança, coalhada, marsupiais, vangloriar, acabrunhado, excepcional, ricochetear, saguões). Em média, os itens apresentaram índices de discriminação de 0,25 ($DP = 0,20$) e de 0,41 ($DP = 0,22$) para os critérios A1 e A2, respectivamente.

Análise dos Itens

Primeiramente, foram selecionados os itens que apresentaram um índice de discriminação (D) maior do que 0,40. Com esse ponto de corte, restaram apenas 20 itens segundo o Critério A1, enquanto que pelo Critério A2 foram 34 itens, que respectivamente compuseram a Lista A1 e a Lista A2. Pelo fato de terem sido mantidos poucos itens após a extração do índice D pelo Critério A1, nenhuma outra exigência foi feita para a Lista A1, tendo sido considerada finalizada a seleção. Para este conjunto de itens, o nível de dificuldade variou de 0,52 a 0,84 ($M = 0,71$; $DP = 0,11$) e o de discriminação variou de 0,41 a 0,84 ($M = 0,53$; $DP = 0,12$), tendo havido um valor adequado de confiabilidade a partir do alfa de Cronbach (0,85). Os itens apresentaram índices satisfatórios de correlação item-total, que variou de 0,27 a 0,60 ($M = 0,43$; $DP = 0,08$). A Tabela 1 (colunas 1 a 4) resume os resultados.

Segundo Pasquali (1996), é desejável que a distribuição dos itens em termos de níveis de dificuldade se aproxime da distribuição normal, com uma menor proporção de itens com níveis extremos de dificuldade. O próximo passo da análise conduzida para seleção de itens para a Lista A2 tenta seguir esta recomendação. Assim, dos 34 itens que apresentaram $D \geq 0,40$, nenhum foi considerado muito difícil ($p < 0,20$), nove itens foram considerados difíceis ($0,20 \leq p < 0,40$), quatro apresentaram nível médio de dificuldade ($0,40 \leq p < 0,60$), 18 foram fáceis ($0,60 \leq p < 0,80$) e três foram muito fáceis ($p \geq 0,80$). Pelo fato de nenhum item ter tido nível de dificuldade muito alto, foram selecionados todos os nove

Tabela 1

Índices de Discriminação (D), Dificuldade (p) e Correlação Item-Total (I-T) para Todas as Palavras Selecionadas de Acordo com o Critério de Acerto do Manual (Lista A1) e pelo Critério Adotado por Lúcio et al. (2009; Lista A2)

Lista A1	D	p	I-T	Lista A2	D	p	I-T
Acabrunhado	0,49	0,73	0,37	Acabrunhado	0,55	0,33	0,46
Acordar	0,41	0,77	0,34	Acordar	0,65	0,65	0,49
Aeronáutica	0,41	0,84	0,45	Aeronáutica	0,78	0,56	0,63
Atlas	0,44	0,78	0,41	Armadura	0,58	0,74	0,51
Atmosfera	0,51	0,81	0,45	Atmosfera	0,71	0,68	0,56
Costas	0,41	0,83	0,42	Bandeja	0,54	0,82	0,52
Durex	0,55	0,82	0,51	Chocalho	0,62	0,66	0,51
Exausto	0,54	0,77	0,51	Coalhada	0,63	0,53	0,49
Excepcional	0,66	0,56	0,46	Costas	0,60	0,75	0,51
Hipócrita	0,66	0,52	0,43	Exausto	0,81	0,60	0,60
Hospedaria	0,41	0,80	0,38	Excepcional	0,78	0,33	0,58
Luxuoso	0,47	0,81	0,43	Guitarra	0,56	0,81	0,55
Marsupiais	0,48	0,70	0,38	Hipócrita	0,53	0,27	0,44
Perseverança	0,45	0,70	0,33	Hospedaria	0,77	0,59	0,58
Repugnante	0,84	0,60	0,60	Luxuoso	0,67	0,60	0,51
Rescindido	0,60	0,65	0,47	Marsupiais	0,66	0,35	0,53
Ricochetear	0,45	0,62	0,27	Perseverança	0,73	0,42	0,53
Saguões	0,57	0,56	0,37	Quiosque	0,62	0,64	0,47
Trouxe	0,49	0,79	0,43	Repugnante	0,67	0,34	0,54
Vangloriar	0,77	0,54	0,55	Rescindido	0,46	0,25	0,41
				Ricochetear	0,44	0,25	0,36
				Saguões	0,55	0,31	0,42
				Tempestade	0,46	0,82	0,47
				Trouxe	0,58	0,71	0,46
				Vangloriar	0,80	0,34	0,58

itens considerados difíceis. De maneira análoga, todos os itens com nível de dificuldade médio e os muito fáceis foram selecionados, uma vez que ocorreram em pequeno número. Nesse sentido, um novo ponto de corte foi aplicado apenas para os itens fáceis, que corresponderam à maioria dos itens com índices adequados de discriminação. Para esses itens, foi aplicado um índice *D* mínimo ainda mais exigente, de 0,56. Com esse

novo ponto de corte, foi possível selecionar nove itens dentro deste nível de dificuldade.

A partir desta análise, restaram 25 itens na Lista A2, cujos níveis de dificuldade variaram de 0,25 a 0,82 ($M = 0,53$; $DP = 0,20$) e os níveis de discriminação entre 0,44 e 0,81 ($M = 0,63$; $DP = 0,11$). O alfa de Cronbach desta lista foi levemente superior ao obtido na Lista A1 (0,90), assim como os índices de correlação item-total

($M = 0,51$; $DP = 0,06$), que variaram de 0,36 a 0,63. Os resultados para a análise dos itens pelo Critério A2 também se encontram na Tabela 1 (colunas 5 a 8).

Estatísticas Descritivas

A Tabela 2 apresenta as estatísticas descritivas dos itens selecionados. Pode-se observar que o efeito de teto reportado por Lúcio et al. (2009) reduz drasticamente para ambas as listas produzidas, apesar de ser menor para a Lista 2, conforme se observa a partir dos valores de assimetria

e curtose. Uma inspeção visual das Figuras 1 a 4 também confirma essa tendência. A distribuição dos escores dos participantes na versão original do subteste é acentuadamente assimétrica à esquerda (Figuras 1 e 2), apesar de esta assimetria ocorrer em menor proporção quando se utiliza o critério de acerto A2 (Figura 2). Após a análise dos itens, a distribuição dos escores fica mais semelhante a uma curva normal para ambas as listas geradas (Figuras 3 e 4), mas o efeito de teto praticamente desaparece apenas na Lista A2 (Figura 4).

Tabela 2

Estatística Descritiva dos Itens de Cada Lista Após a Seleção Segundo os Critérios A1 e A2

Lista/ano	<i>N</i>	Média	<i>DP</i>	Mínimo	Máximo	Assimetria	Curtose
A1 (20 itens)							
2º	77	11,2	4,3	0,0	20,0	-0,3	-0,4
3º	83	13,6	4,2	4,0	20,0	-0,4	-0,9
4º	99	14,9	3,9	4,0	20,0	-0,7	-0,2
5º	82	16,7	3,7	4,0	20,0	-1,4	1,7
Total	341	14,1	4,5	0,0	20,0	-0,6	-0,4
A2 (25 itens)							
2º	77	8,0	5,4	0,0	23,0	0,4	-0,5
3º	83	12,4	5,8	0,0	23,0	-0,1	-0,9
4º	99	14,7	5,5	2,0	25,0	-0,1	-0,6
5º	82	17,8	5,2	1,0	25,0	-0,9	0,5
Total	341	13,3	6,4	0,0	25,0	-0,2	-0,9

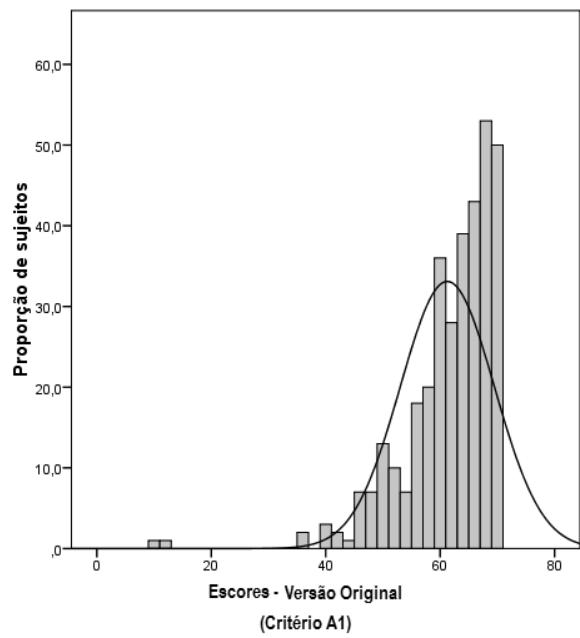


Figura 1. Distribuição dos escores dos participantes da amostra geral na versão original do subteste (Critério A1).

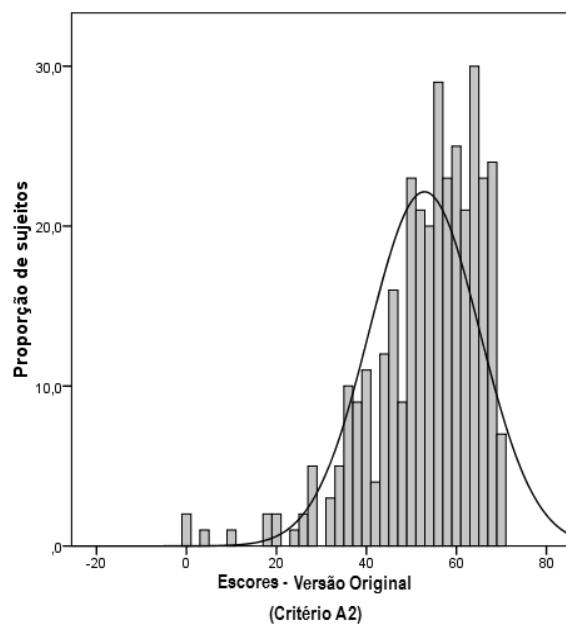


Figura 2. Distribuição dos escores dos participantes da amostra geral na versão original do subteste (Critério A2).

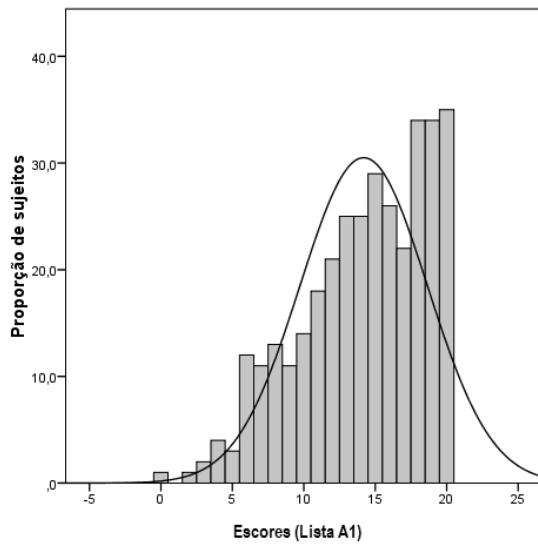


Figura 3. Distribuição dos escores dos participantes da amostra geral após a análise dos itens do subteste (Critério A1).

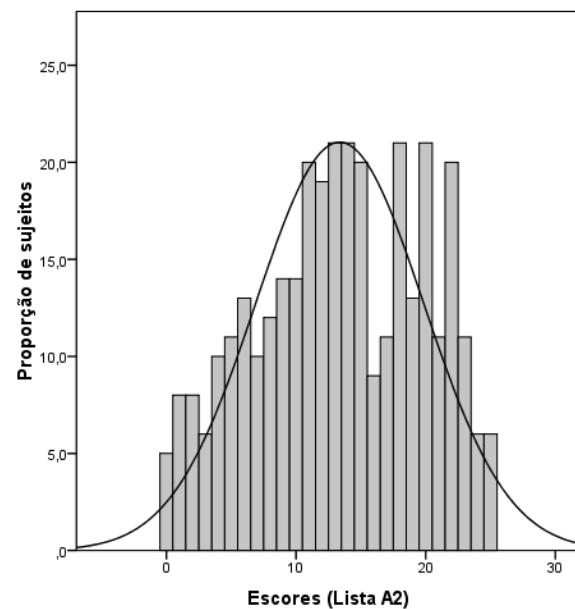


Figura 4. Distribuição dos escores dos participantes da amostra geral após a análise dos itens do subteste (Critério A2).

Os percentis obtidos a partir dos escores dos participantes do estudo podem ser vistos na Tabela 3. Observa-se que faz sentido, após a análise dos itens, acrescentar uma nova faixa de percentil para além do apresentado na versão original do teste, a saber, o percentil 5. Assim,

as listas de palavras passam a contar com cinco faixas percentílicas, variando de P5 a P99. Apenas no 5º ano na Lista A1 não aparece o percentil 99, ocorrendo um efeito de teto no percentil 75. Nota-se que o número de itens presentes em cada uma das Listas é diferente.

Tabela 3

Percentis Obtidos para Todos os Anos Escolares e para a Amostra Geral, a Partir dos Escores Apresentados nas Listas A1 e A2

Ano	Lista A1 (20 itens)					Lista A2 (25 itens)				
	P5	P25	P50	P75	P99	P5	P25	P50	P75	P99
2º	3,0	8,0	12,0	14,0	20,0	0,0	4,0	8,0	12,0	23,0
3º	6,0	10,0	15,0	17,0	20,0	3,0	7,0	13,0	17,0	23,0
4º	7,0	12,0	15,0	18,0	20,0	4,0	11,0	14,0	19,0	25,0
5º	8,1	15,0	18,0	20,0	20,0	8,0	14,0	20,0	22,0	25,0
Total	6,0	11,0	15,0	18,0	20,0	2,0	9,0	13,0	19,0	25,0

Análise Qualitativa

Para ambas as listas geradas a partir da análise dos itens, houve uma predominância de palavras de baixa frequência de ocorrência (95% para a Lista A1 e 96% para a Lista A2), de tamanho grande (90% dos itens com mais de seis letras na Lista A1 e 100% na A2) e regulares, dependentes ou não de contexto (75% das na Lista A1 e 72% na Lista A2). Análises univariadas de variância mostraram que os índices psicométricos dos itens não se diferiram em função da classificação de regularidade das palavras nem se considerando o efeito geral (índice D : $F(2,19) = 0,570, p = 0,0576$; precisão: $F(2,19) = 3,020, p = 0,075$; correlação item-total: $F(2,19) = 0,454, p = 0,643$), nem para a análise de *post-hoc* no teste de Bonferroni (menor valor de $p = 0,09$).

Discussão

A presente pesquisa conduziu uma análise clássica dos itens do subteste de leitura do TDE a partir de uma amostra de escolares de 2º ao 5º ano da cidade de Belo Horizonte. Considerando-se as questões levantadas por Lúcio et al. (2009), a análise foi feita a partir de dois critérios de acerto, um apresentado pelo manual do teste e outro proposto pelas autoras. De uma maneira geral, a análise aqui reportada confirma as questões levantadas naquele trabalho a respeito dos índices de dificuldade dos itens do subteste, assim como das considerações feitas sobre os problemas do critério de acerto utilizado no manual.

O questionamento colocado pelas autoras de que o subteste, em sua forma original, teria problemas em discriminar as habilidades de leitores mais competentes ou dos mais experientes foi confirmado a partir da análise dos índices de discriminação e de dificuldade dos itens. No primeiro caso, foi observado que apenas uma pequena quantidade de itens que apresentou índice $D > 0,40$ (20 pelo Critério A1 e 34 pelo Critério A2), o que indica que a maioria dos itens presentes no subteste não é capaz de diferenciar com precisão as habilidades pretendamente medidas por ele. Isso fica nítido quando se observa a quantidade de itens que apresentaram índice D próximo de zero, podendo-se distinguir 20 palavras que atingiram índice D menor de 0,1 pelo Critério A1 (pato, mato, vela, fita, janela, minha, saco, caju, sapato, agulha, caminhão, agora, tijolo, tamanho, arte, isca, moeda, palavra, querido, campo) e quatro pelo Critério A2 (pato, mato, fita, caju).

No segundo caso, que diz respeito à dificuldade (p), pode-se verificar uma grande quantidade de itens que apresentaram proporção de acerto maior do que 60%. Pelo Critério A1, foram 66 itens, e pelo A2 foram 57 (lembrando que o subteste apresenta 70 itens). Uma alta proporção de acertos parece ser uma característica das tarefas de leitura em voz alta de palavras isoladas do português (Lúcio, Moura, Nascimento, & Pinheiro, 2012; Pinheiro, 2011). De fato, em vários estudos que fazem uso de listas experimentais de palavras, a média de acerto na leitura gira entre 70% e 80% de acerto (p. ex., Justi & Justi, 2009;

Lúcio, Pinheiro, & Nascimento, 2010; Pinheiro, Lúcio, & Silva, 2008; Salles & Parente, 2002). Isso demonstra a importância da análise de itens, uma maneira mais segura de se impedir que uma grande quantidade de itens fáceis ou muito fáceis faça parte do teste, evitando, por consequência, a presença de um efeito de teto e a entrada de itens redundantes que não acrescentam informações para a avaliação das diferenças individuais³.

As diferenças anteriormente apresentadas sobre a distribuição dos índices de dificuldade e de discriminação a partir dos diferentes critérios de acerto utilizados também confirmam a hipótese de que a consideração de respostas de silabação e de correção espontânea como acerto contribui para a diminuição da variabilidade dos escores. Todos os dados aqui coletados confirmaram que a configuração dos escores pelo Critério A2 foi mais informativa das habilidades dos sujeitos do que pelo critério exposto no manual. Sendo assim, pudemos observar que o Critério A2 produziu itens com melhores índices de discriminação e com níveis de dificuldade cobrindo faixas mais amplas. Isso repercutiu na produção de uma maior quantidade de itens disponíveis para compor as listas após a análise (20 itens pelo Critério A1 e 25 pelo Critério A2). Além disso, os valores de consistência interna foram mais altos para A2 do que para A1. Pela análise

dos gráficos apresentados nas Figuras 1 a 4, e dos valores de assimetria e curtose presentes na Tabela 2, observou-se também que a distribuição dos escores se assemelhou mais a uma curva normal apenas quando adotado o Critério A2. Já a análise da distribuição dos percentis mostrou que pelo Critério A2 é possível distinguir o percentil 99 em todos os anos escolares (inexistindo no 5º ano pelo Critério A1) e que escores máximos são obtidos nesta lista apenas entre os leitores mais experientes (a partir do 4º ano).

Finalmente, em ambas as listas, observou-se uma tendência à seleção de itens de baixa frequência, grandes e regulares (ou dependentes de regras contextuais). Utilizando uma lista de palavras de baixa frequência, Lúcio et al. (2012) também demonstraram que índices psicométricos foram mais adequados para as palavras grandes, mas, ao contrário do presente estudo, as palavras irregulares foram em média mais difíceis. Nos itens do TDE, os índices psicométricos não se diferiram em função da classificação de regularidade das palavras. Desse modo, a configuração final das listas caracterizou-se pela ausência de emparelhamento das propriedades psicolinguísticas dos itens (Pinheiro & Rothe-Neves, 2001). Isso inviabiliza o estabelecimento de uma análise dos efeitos que são comumente relatados na literatura para a avaliação do funcionamento do processamento cognitivo da leitura (p. ex., Coltheart et al., 2001), tal como se observa na versão original do teste (ver nota de rodapé nº 1).

³ Obviamente, dependendo dos objetivos que se pretende com os escores do teste, seria possível inserir itens com níveis de dificuldade mais baixos, por exemplo, quando o interesse está na criação de testes neuropsicológicos ou de triagem. Nestes casos, a seleção dos itens dá-se basicamente pelo nível de discriminação, o que permite que o teste seja capaz de discernir entre aqueles que têm a habilidade (independente de qual nível) e aqueles que não a possuem. A questão que se coloca é que o subteste foi criado com o objetivo de avaliar as capacidades básicas de leitura, tendo por amostra normativa crianças com desenvolvimento normal cursando os anos iniciais do Ensino Fundamental, o que o exclui da categoria de triagem. Desta forma, é necessário considerar a proporção de itens em termos de nível de dificuldade que entrarão no teste, da mesma forma que é preciso considerar os índices de discriminação, conforme se estabeleceu aqui a partir das recomendações de Pasquali (1996).

Considerações Finais

A presente pesquisa indicou a importância da análise e seleção empírica dos itens no momento de se criar testes com o objetivo de investigar as diferenças individuais nas habilidades por eles avaliadas. A análise de itens é uma etapa fundamental na construção de testes e nunca deve ser negligenciada. Os dados, aqui produzidos para o subteste de leitura do TDE, indicaram nitidamente a mudança de um padrão de efeito de teto da distribuição dos escores para uma distribuição mais próxima à da curva normal.

Também ficou demonstrado que a teoria tem um papel decisivo na qualidade final dos

itens selecionados. Desse modo, a teoria de reconhecimento de palavras, tal como concebida pelos modelos de dupla-rota (Coltheart et al., 2001), faz uma diferenciação entre o reconhecimento por meio de dois processos, o fonológico e o ortográfico. O primeiro processo faz uso dos sons da linguagem falada na aprendizagem sobre como decodificar a linguagem escrita (Wagner & Torgesen, 1987) e o segundo faz uso sobre a ortografia de uma língua tanto no processamento da linguagem falada quanto da escrita (Wagner & Baker, 1994). No processo de decodificação fonológica, pelo menos nas ortografias mais transparentes do que a do inglês ou francês, por exemplo, as palavras são reconhecidas por meio da aplicação de regras de correspondência grafema-fonema (Protopapas & Vlahou, 2009) e no ortográfico, as representações lexicais, semânticas e fonológicas da palavra são recuperadas como um todo, de uma maneira automática. As tarefas de leitura em voz alta de palavras, principalmente as irregulares, constituem uma das formas de se avaliar o reconhecimento ortográfico de palavras. Para a avaliação do processo de decodificação fonológica, são mais adequadas tarefas de leitura de pseudopalavras.

Nesse sentido, o critério de acerto adotado em um teste deve estar de acordo com o que é esperado pela teoria que o embasa e pelos propósitos da testagem. Por esta razão, em nossa análise do subteste de leitura do TDE, encontramos diferenças importantes entre os resultados produzidos pelos Critérios A1 e A2, apesar de a seleção final de itens ter melhorado a distribuição dos escores em ambos os casos. No entanto, por não se ter controlado as variáveis psicolinguísticas do teste (ver análise qualitativa, na seção resultados), não se prestará a um diagnóstico sobre os processos de leitura, mas com as mudanças aqui introduzidas poderá ser um teste adequado para uma primeira estimativa para esta habilidade, de uma maneira geral.

Recomendações Futuras

No momento de se criar listas de palavras com o objetivo de avaliar o desempenho cognitivo na leitura, particularmente quando o interesse

é investigar o reconhecimento de palavras, é necessário considerar dois aspectos distintos: um teórico e outro psicométrico. O aspecto teórico guiará a escolha das características psicolinguísticas dos itens e a determinação do critério de acerto do teste. O ponto de vista psicométrico, dentro de uma perspectiva clássica, determinará a distribuição dos itens em função dos índices de dificuldade e de discriminação por eles apresentados. As listas de palavras aqui geradas a partir da versão original do subteste de leitura do TDE seguem apenas o último ponto de vista. Nesse sentido, recomendamos que elas sejam usadas apenas para avaliar as habilidades de leitura de palavras isoladas esperadas para cada faixa de ano escolar investigada, tal como estabelecido em sua forma original.

Referências

- American Educational Research Association, American Psychiatry Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. New York: American Educational Research Association.
- Associação Brasileira de Empresas de Pesquisa. (2003). *Critério de classificação econômica Brasil*. Recuperado em 15 de dezembro, 2011, de <http://www.marketanalysis.com.br/arquivos-download/biblioteca/cceb-1.pdf>
- 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*. Recuperado em 22 de dezembro, 2011, de <http://www.pol.org.br>
- Flynn, J. R. (2006). O efeito Flynn: Repensando a inteligência e aquilo que a afeta. In C. Flores-Mendoz, R. Colom, A. Andrés-Pueyo, R. J. Haier, & S. A. Petrill, *Introdução à Psicologia das Diferenças Individuais* (pp. 387-411). Porto Alegre, RS: ArtMed.
- 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.

- Lúcio, P. S. (2008). *Investigação psicométrica de uma tarefa de leitura em voz alta de palavras isoladas* (Dissertação de mestrado, Programa de Pós-Graduação em Psicologia, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil).
- Lúcio, P. S., Moura, R. J., Nascimento, E., & Pinheiro, A. M. V. (2012). Construção de uma tarefa de leitura em voz alta de palavras: Análise psicométrica dos itens. *Psicologia: Reflexão e crítica*, 25(4), 662-670.
- Lúcio, P. S., Pinheiro, A. M. V., & Nascimento, E. (2009). O impacto da mudança do 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.
- Pasquali, L. (1996). *Teorias e métodos de medida em ciência do comportamento*. Brasília, DF: Laboratório de Medidas em Ciência do Comportamento, Instituto de Psicologia, Universidade de Brasília.
- Pinheiro, A. M. V. (1996). *Contagem de frequência de ocorrência de palavras expostas a crianças da 1ª à 4ª série do Ensino Fundamental*. São Paulo, SP: Associação Brasileira de Dislexia.
- Pinheiro, A. M. V. (2007). Anexo 2. In I. Sim-Sim & F. L. Vianna (2007), *Para a avaliação do desempenho de leitura* (pp. 121-131). Lisboa, Portugal: Gabinete de Estatística e Planeamento da Educação. (Original publicado em 1996). Recuperado em 10 de Setembro, 2010, de <http://www.gepe.min-edu.pt>
- Pinheiro, A. M. V. (2011). Transparência ortográfica e o efeito de retroalimentação fonológico grafêmico: Implicações para a construção de provas de reconhecimento de palavras. In L. M. Alves, R. Mousinho, & S. A. Capellini (Eds.), *Dislexia, novos tema, novas perspectivas* (pp. 131-151). Rio de Janeiro, RJ: Wak.
- Pinheiro, A. M. V., Lúcio, P. S., & Silva, D. M. R. (2008). Avaliação cognitiva de leitura: O efeito de regularidade grafema-fonema e fonema-grafema na leitura em voz alta de palavras isoladas. *Psicologia: Teoria e Prática*, 10(2), 16-30.
- Pinheiro, A. M. V., & Rothe-Neves, R. (2001). Avaliação cognitiva de leitura e escrita: As tarefas de leitura em voz alta e ditado. *Psicologia: Reflexão e Crítica*, 14(2), 399-408.
- Protopapas, A., & Vlahou, E. L. (2009). A comparative quantitative analysis of Greek orthographic transparency, *Behavior Research Methods*, 41(4), 991-1008.
- Raven, J., Raven, J. C., & Court, J. H. (1993). *Manual das Matrizes Progressivas de Raven: Escala especial* (L. A. Angelini, I. C. B. Alves, E. M. Custódio, W. F. Duarte, & J. L. M. Duarte, Trad. e adaptação). São Paulo, SP: Centro Editor de Testes e Pesquisas em Psicologia.
- Salles, J. F., & Parente, M. A. M. (2002). Processos cognitivos na leitura de palavras em crianças: Relações com compreensão e tempo de leitura. *Psicologia: Reflexão e Crítica*, 15(2), 321-331.
- Seidenberg, M. S. (2006). Connectionist Models of Reading. In G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 235-25). Oxford, UK: Oxford University Press.
- 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.
- Wagner, R. K., & Barker, T. A. (1994). The development of orthographic processing ability. In V. W. Berninger (Ed.), *The variety of orthographic knowledge 1: Theoretical and developmental issues* (pp. 243-276). Dordrecht, Netherlands: Kluwer Academic.
- Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, 101, 192-212.
- Wechsler, D. (1991). *WISC-III: Escala de Inteligência Wechsler para Crianças: Manual* (3. ed., V. L. M. de Figueiredo, Adaptação e padronização). São Paulo, SP: Casa do Psicólogo.

Recebido: 02/07/2013

1ª revisão: 08/08/2013

Aceite final: 14/08/2013