

O IMPACTO DA MUDANÇA NO CRITÉRIO DE ACERTO NA DISTRIBUIÇÃO DOS ESCORES DO SUBTESTE DE LEITURA DO TESTE DE DESEMPENHO ESCOLAR¹

Patrícia Silva Lúcio^{*}
Ângela M. V. Pinheiro[#]
Elizabeth do Nascimento[¶]

RESUMO. O estudo investiga o impacto da introdução de uma nova classe de erros no subteste de leitura do Teste de Desempenho Escolar na distribuição dos escores. Participaram do estudo 306 crianças de 1ª a 4ª séries da rede de ensino de Belo Horizonte. Para cada participante foram conferidos dois escores: um estabelecido a partir de critério do manual (EB1) e o outro a partir do critério adotado no presente estudo (EB2) – que considerou incorretas as respostas de silabação e de correção espontânea. Os resultados mostraram que a adoção do critério EB2 tornou o teste mais discriminativo, mas não impediu o aparecimento de um efeito de teto, nem proporcionou uma distribuição normal dos escores. As análises mostraram que, além do critério permissivo, o excesso de palavras fáceis e a escassez de palavras difíceis prejudicaram a variabilidade dos escores. Hipóteses sobre a dificuldade das palavras são discutidas à luz da psicolinguística.

Palavras-chave: reconhecimento de palavras, critérios de acertos, poder de discriminação de itens.

THE IMPACT OF THE CHANGE IN THE CRITERION OF SCORING CORRECT RESPONSES IN THE DISTRIBUTION OF THE SCORES OF THE READING SUBTEST OF PERFORMANCE SCHOOL TEST

ABSTRACT. The study investigates the impact of the introduction of a new class of errors within the original error set of the reading subtest of the Performance School Test on the distribution of the scores. The sample consisted of 306 1st - 4th grades of the school system of Belo Horizonte who participated in the study. For each child two scores were given: one based on the manual criteria (EB1) and the other on the criteria introduced in the present study (EB2) – which considered explicit sounding out of syllables or of part of the stimuli and self-corrections as incorrect responses. The results showed that the choice of the EB2 made the test more discriminative, but such a modification did not prevent the emergence of a ceiling effect, nor did it enable a normal distribution of the scores. The analysis showed that apart from the permissive criteria the excess of easy words and the shortage of difficult words affected the variability of the scores. Hypotheses about the difficulty of words were discussed in light of the psycholinguistics.

Key words: Word recognition task, criteria for correct responses, item discrimination power.

EL INFLUJO DEL CAMBIO DE CRITERIO DE ACIERTO EN LA DISTRIBUCIÓN DE LOS PUNTAJES DE LA SUBPRUEBA DE LECTURA EN LA PRUEBA DE RENDIMIENTO ESCOLAR

RESUMEN. El estudio investiga el influjo que la introducción de una nueva clase de errores en la subprueba de lectura de la Prueba de Rendimiento Escolar tuvo en la distribución de los puntajes de la referida subprueba. Participaron de la investigación 306 niños de 1ª a 4ª clase de la Red de Enseñanza de Belo Horizonte. Para cada participante han sido conferidos

¹ Apoio: FAPEMIG e do CNPq.

^{*} Mestre em Psicologia do Desenvolvimento pela UFMG. Professora do Curso de Psicologia da Faculdade de Minas – FAMINAS/Muriaé.

[#] Doutora em Psicologia Cognitiva pela Universidade de Dundee (Escócia). Professora Titular do Departamento de Psicologia da Universidade Federal de Minas Gerais-UFMG.

[¶] Doutora em Psicologia pela UNB. Professora Adjunta do Departamento de Psicologia da UFMG.

dos puntajes: uno establecido a partir de criterio del manual (EB1) y otro a partir del criterio adoptado en la presente investigación (EB2) – que consideró incorrectas las respuestas de silabación y de corrección espontánea. Los resultados han mostrado que la adopción del criterio EB2 tornó la prueba más discriminadora, pero esa modificación no impidió el asomo de un efecto de techo, ni proporcionó una distribución normal de los puntajes. Los análisis han mostrado que, además del criterio permisivo, el exceso de palabras fáciles y la escasez de palabras difíciles perjudicaron la variabilidad de los puntajes. Hipótesis sobre la dificultad de las palabras son discutidas a la luz de la psicolingüística.

Palabras-clave: Reconocimiento de palabras, Criterios de aciertos; Poder de discriminación de ítems.

A tarefa de leitura em voz alta de palavras isoladas é reconhecida em diversas línguas alfabéticas como um método eficaz de avaliação cognitiva da leitura, tendo se mostrado bastante eficiente como forma de testar os componentes do modelo de dupla rota, cuja versão mais recente é a de Coltheart, Rastle, Perry, Langdon e Ziegler (2001). Tal modelo, que foi desenvolvido no contexto da língua inglesa e cuja aplicação para o português foi testada por Pinheiro (1995), pressupõe que a leitura pode ocorrer por meio de dois processos distintos: um processo fonológico, em que a pronúncia da palavra é construída por meio de decodificação fonológica – que consiste na aplicação de regras que convertem a sequência grafemas que compõem a palavra escrita em uma sequência de fonemas, com subsequente acesso do significado do estímulo lido; e um processo lexical, em que tanto o significado quanto a pronúncia da palavra são recuperados após o reconhecimento de sua grafia por meio de um mecanismo visual. Acredita-se que o processo fonológico (também conhecido como processo de mediação fonológica) seja utilizado por leitores iniciantes, que, gradualmente, vão formando um vocabulário de visão (um dicionário interno – ou léxico – que contém todas as palavras que reconhecemos de forma visual), passando, assim, a reconhecer lexicalmente as palavras (para uma revisão sobre as evidências a favor da existência desses diferentes processos, ver Coltheart et al., 2001).

No Brasil, a avaliação cognitiva da leitura tem sido feita principalmente com utilização de listas de palavras que variam em termos de características psicolingüísticas contrastantes (como, regularidade, comprimento e frequência) que, quase em sua totalidade, não passaram pelos processos psicométricos de validação, de precisão e de criação de normas (por exemplo, Pinheiro, 1995; Salles, 2005). O único teste de leitura em voz alta de palavras isoladas que apresenta normas para a população brasileira, além de evidências psicométricas de validade e precisão, é o subteste de leitura do Teste de Desempenho Escolar – TDE (Stein, 1994). Segundo informações do manual,

este subteste procura avaliar a habilidade básica de decodificação, que, de acordo com a literatura (p. ex., Stuart & Coltheart, 1988; Ehri, 1992), é uma etapa necessária para a aquisição de um vocabulário de visão, cujo desenvolvimento e utilização automática para o reconhecimento de palavras assinalam a leitura competente.

Não obstante, ao considerar os critérios de acerto do referido subteste, nota-se a falta de menção ao tratamento que se deve dar às respostas em que a leitura é feita de forma explicitamente silabada (que é caracterizada pela presença de pausas entre as sílabas, ou mesmo entre as letras²), ficando-se, então, com a impressão de que esse tipo de resposta deva ser considerado correto³. Acreditamos que, embora a leitura silabada seja um dos expoentes do processo de decodificação fonológica, é, sobretudo, uma resposta que denota que esse processo está sendo efetuado de forma linear e sem o mínimo do automatismo que se espera na leitura, mesmo das crianças nas fases iniciais de aquisição da leitura.

Assim, a adoção de um critério de acerto tão permisivo no subteste em questão apresenta pelo menos duas repercussões práticas. A primeira delas é que a desconsideração de uma leitura explicitamente silabada significa ignorar o fato de que uma decodificação lenta e laboriosa por crianças, mesmo na 1ª e 2ª séries, limita os recursos de processamento que são normalmente direcionados para a compreensão. Além disso, vale lembrar que a língua portuguesa, por apresentar apenas dois casos de irregularidade para a leitura (Parente, Silveira, & Lecours, 1997; Scliar-Cabral, 2003), não impõe muitas dificuldades àqueles que

² A pausa entre as letras ocorre quando a sílaba é composta apenas de uma letra (por exemplo, ler a palavra *aceso* como “a-ce-so”, o que é diferente da pura nomeação de letras na qual a criança simplesmente soletra a palavra, ou seja, não realiza a leitura.

³ De fato, em comunicação pessoal, muitos usuários do subteste de leitura do TDE nos informaram que consideraram como acerto as respostas em que a leitura foi feita de forma explicitamente silabada.

leem por mediação fonológica. Assim, apesar de serem esperadas diferenças na habilidade de decodificação entre as crianças mais novas e as mais velhas⁴, essas diferenças podem não ser muito grandes, pelo menos a partir da 2ª série. Por outro lado, mais grave ainda é desconsiderar a leitura explicitamente silabada nas respostas das crianças a partir da 3ª série, uma vez que, nos estágios mais avançados do desenvolvimento da leitura, as palavras tendem a ser reconhecidas por meio do processamento lexical, com base na especificação da representação ortográfica, semântica e fonológica das palavras previamente arquivadas na memória lexical, e não pelo processo inicial em que o significado da palavra é obtido por meio de mediação fonológica. Conseqüentemente, não quantificar os sinais de leitura fonológica nesse estágio significa que uma importante etapa do desenvolvimento da habilidade da leitura não será captada pelo teste, o que dificulta a avaliação das diferenças entre os leitores iniciantes e os mais experientes (ou mesmo entre os mais experientes entre si).

A segunda repercussão é que, como as crianças, mesmo as que estão no início da alfabetização, conseguem ler as palavras de maneira eficiente por meio do processo fonológico, principalmente se forem regulares, os níveis de acerto no teste serão elevados mesmo entre os leitores iniciantes, o que pode gerar um efeito de teto no teste, prejudicando a avaliação das diferenças individuais entre os sujeitos superiores na habilidade de leitura entre si (ou seja, aqueles que se encontram posicionados na extremidade direita de uma curva normal). O efeito de teto também prejudica a avaliação das diferenças individuais entre os sujeitos menos capazes, uma vez que uma grande proporção de acertos no teste é exigida para que eles possam ser enquadrados na classificação inferior (percentil 25 ou menos), fazendo com que indivíduos que obtiveram uma alta proporção de acertos tenham a mesma classificação descritiva que aqueles que obtiveram

um escore baixo ou muito baixo. Isso foi demonstrado, por exemplo, por Capellini, Mendes e Salgado (2003), que, trabalhando com crianças de 1ª a 4ª séries com problemas de audição, mostraram que mais de 90% da amostra apresentaram desempenho abaixo do esperado pela série no subteste de leitura. Segundo os autores, o alto ponto de corte do teste para a classificação inferior fez com que alunos que obtiveram um percentual de acertos muito baixo ficassem no mesmo grupo de outros que apresentaram resultados mais altos, o que dificultou a discriminação das habilidades dos participantes.

Em suma, a adoção de um critério permissivo de acerto, aliada a uma língua relativamente consistente, que pode ser lida sem grandes problemas pelo uso de mediação fonológica, fará com que os níveis de acerto no teste seja elevado entre os leitores de todos os níveis (competentes ou não competentes, iniciantes ou experientes). Isso prejudica a variabilidade dos escores e a avaliação das diferenças individuais na habilidade investigada. Tendo em vista essas questões, o presente trabalho avaliou o desempenho de uma amostra de crianças da 1ª à 4ª séries do Ensino Fundamental da cidade de Belo Horizonte no subteste de leitura do TDE. Para tanto foram utilizados dois critérios de acerto. Um deles é apresentado no manual, e no outro, em adição às respostas consideradas incorretas no manual, foram também consideradas como incorretas respostas de silabação e de correção espontânea (ou seja, quando a criança emite uma resposta considerada incorreta e, em seguida, corrige espontaneamente seu próprio erro). Os objetivos foram descrever o desempenho dos participantes em ambos os critérios, comparando-se os resultados com os da amostra de padronização do TDE, e avaliar o impacto da adoção de um critério de acerto e do outro na distribuição dos escores.

MÉTODO

Procedimentos éticos

A presente pesquisa faz parte de um projeto maior, intitulado “Avaliação de Competências Psicológicas da População Infanto-Juvenil de Belo Horizonte: Inteligência e Habilidades de Leitura e Escrita” e aprovado pelo Comitê de Ética em Pesquisas com Seres Humanos – COEP, da UFMG em 09/03/2005 (processo n.º ETIC 347/04).

⁴ Ao mesmo tempo em que, com o avanço da escolaridade, nota-se um abandono do processo fonológico como recurso primordial no reconhecimento de palavras em favor de um reconhecimento pelo processo lexical, o processo fonológico que se mantém disponível torna-se totalmente automatizado, o que permite a leitura rápida e eficaz tanto de palavras pouco frequentes quanto de pseudopalavras. Assim, o processo fonológico é mais eficiente nos leitores mais experientes do que nos mais jovens (como foi demonstrado por Guimarães (2003) em estudo com crianças brasileiras).

Participantes

Participaram do estudo 306 crianças da 1ª à 4ª série, de ambos os sexos (52,6% do sexo feminino), com média de idade de 8 anos e 9 meses, vindas de seis escolas públicas (duas municipais e quatro estaduais, N = 259) e de uma escola particular (N = 47) da rede de ensino de Belo Horizonte.

Instrumentos

Além do subteste de leitura do TDE, os participantes responderam a uma prova computadorizada de leitura e a testes de inteligência (Raven e subtestes verbais do WISC-III). Somente serão avaliados os resultados no subteste de leitura do TDE.

Procedimento

As crianças foram testadas individualmente, em uma sala especialmente fornecida pela escola para o estudo. As provas foram aplicadas por uma equipe de estudantes do 3º ano de graduação em psicologia, incluindo dois bolsistas PROBIC/FAPEMIG, 14 alunos de voluntários e a primeira autora (à época, estudante de mestrado). A equipe foi treinada na aplicação e correção do teste pela própria mestranda em três sessões com duração de duas horas cada. Os aplicadores foram instruídos a marcar conforme as indicações do manual do TDE as palavras consideradas incorretas pelo critério de Stein (1994) e marcar com um círculo as palavras consideradas incorretas pelo nosso critério, ou seja, as respostas de silabação explícita e de correção espontânea. Isso se justifica porque, como dissemos, o reconhecimento visual competente de palavras se caracteriza pela recuperação da pronúncia da palavra por meio de acesso direto à memória, com um mínimo de mediação fonológica. As situações em que a resposta silabada é seguida de correção espontânea nos mostram que a rota fonológica produziu uma resposta antes da rota lexical, o que, a partir das fases iniciais da aquisição da leitura e do tipo de palavra que está sendo lida, é um sinal de um processamento lento (ineficiente) na rota lexical⁵.

⁵ É importante ressaltar que as respostas de correção espontânea representam um tipo de erro mais “evoluído” do ponto de vista do desenvolvimento da leitura do que a emissão de uma resposta de silabação (ou mesmo outros tipos de erro) sem a sua posterior correção. Essas respostas nos informam que, apesar de as representações lexicais do leitor para os itens espontaneamente corrigidos estarem fracas, por a pronúncia desses itens não ter sido recuperada

Assim, os participantes receberam dois escores brutos: um escore segundo os critérios de Stein, que chamamos aqui de EB1, e um escore bruto de acordo com o critério utilizado por nós, nomeado EB2. Os dados foram analisados com o pacote estatístico SPSS 15.0 (2006).

Procedimento de análise estatística

Foi conduzida uma análise para verificar a normalidade nos dados por meio do teste Kolmogorov-Smirnov (K-S). Os dados da amostra do presente estudo não apresentaram distribuição normal nem pelo EB1 (K-S = 2,3; $p < 0,0001$) nem pelo EB2 (K-S = 1,741; $p < 0,05$), o que levou a adotarmos a estatística não-paramétrica para a extração dos resultados (teste T de Wilcoxon). Serão considerados significativos todos os resultados de análises estatísticas que apresentarem um valor de $p < 0,05$.

RESULTADOS

A Tabela 1 mostra as médias de acerto e os desvios-padrão no subteste de leitura do TDE na amostra de padronização e na amostra do presente estudo, segundo os critérios de Stein (1994) (EB1) e os critérios aqui adotados (EB2). Pela análise desta Tabela, pode-se observar que os valores das médias do EB1 são muito semelhantes aos apresentados pela amostra de padronização, com exceção da 1ª série, que em nosso estudo apresentou médias de acerto bem mais elevadas. Outra diferença é que os desvios-padrão das séries iniciais (1ª e 2ª séries) aqui apresentados são bem menores do que os da amostra de padronização; já o EB2, apesar de também ter apresentado na 1ª maiores valores do que os da mesma série na amostra de padronização, apresentou médias de acertos bem inferiores em todas as séries, demonstrando o impacto do critério de correção nos escores dos participantes, conforme se verá.

automaticamente, elas estão em processo de estabelecimento ou de consolidação. Isso se deve ao fato de a recuperação das palavras, mesmo que tardia, requerer algum nível de ativação no léxico mental. Entretanto, apesar desse indício de reconhecimento lexical, é necessário segregar as respostas de correção espontânea daquelas emitidas sem hesitação, uma vez que estas últimas indicam que o leitor já adquiriu a representação completa da palavra, com sua recuperação facilmente acessada do léxico mental.

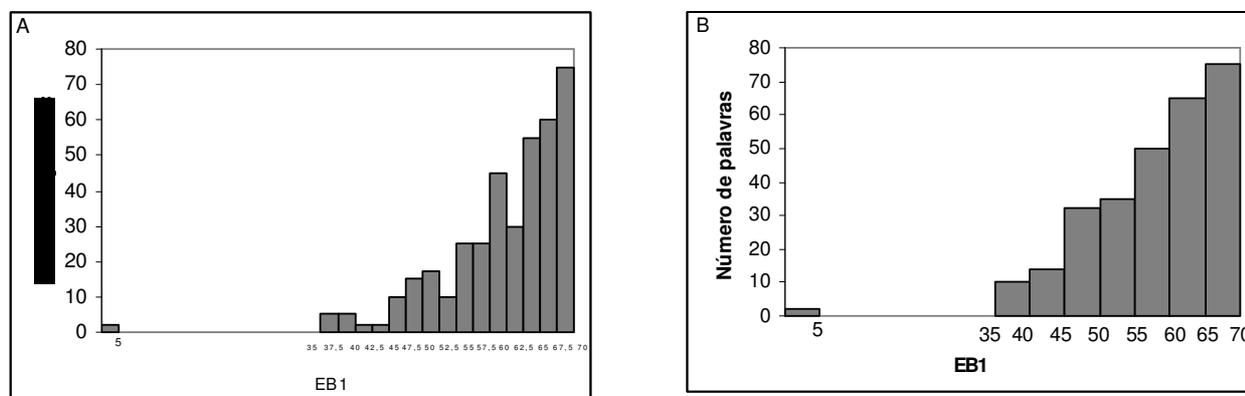
Tabela 1. Média (M) e Desvio Padrão (DP) dos Escores de Acerto dos Participantes para Cada SÉRIE segundo a Amostra de Padronização, do Critério de Stein (EB1) e do Presente Estudo (EB2) e Valores de Mediana (MD) para Ambos os Critérios

Escore bruto/ série	Padronização		EB1		EB2			
	M	DP	M	DP	MD	M	DP	MD
1ª	34,67	27,18	56,40	9,56	59	43,82	13,33	45
2ª	58,83	13,43	60,13	7,16	62	50,65	11,51	53
3ª	64,57	4,67	62,69	5,87	63,5	56,83	7,99	56,5
4ª	66,67	2,56	65,82	4,91	68	61,46	6,63	64
Total	-	-	61,34	7,73	63	53,37	11,86	56

As medianas para todas as séries e ambos os critérios de acertos também podem ser vistos na Tabela 1. O teste T de Wilcoxon para duas amostras relacionadas foi utilizado para a comparação dos escores e mostrou que há diferenças significativas entre o EB1 e o EB2 ($z = 14,011$, $p < 0,0001$), com maiores valores de mediana para o EB1, o que implica que os participantes apresentaram escores significativamente mais elevados no critério do manual do que naquele adotado por esta pesquisa. O próximo passo foi verificar se a mudança no critério apresentou repercussões na forma de distribuição dos dados, ou seja, se esta mudança provocou um efeito na discriminação do teste entre os leitores mais ou menos capazes.

A não-normalidade dos dados parece ter sido devida principalmente à ausência de itens mais difíceis no subteste (efeito de teto), os quais poderiam discriminar de forma

mais eficiente as habilidades dos leitores experientes ou dos mais competentes⁶. Talvez isso tenha feito com que a mudança no critério de acerto não repercutisse em uma distribuição normal dos dados. No entanto, por meio de uma inspeção visual nos histogramas apresentados na Figura 1, podemos observar que uma grande proporção de participantes obteve escores de 60 ou mais pelo EB1, enquanto pelo EB2 a proporção de participantes se concentra em escores maiores que 40 (lembrando que o teste apresenta 70 itens), o que demonstra uma maior variabilidade (ou dispersão em torno da média) de EB2 em relação a EB1. Essa maior variabilidade também fica evidente pela presença de um efeito de teto menor no critério adotado em EB2 (representada por uma menor proporção de indivíduos que acertaram o teste todo ou quase todo em relação ao critério EB1).

**Figura 1.** Histogramas da Distribuição dos Escores a Partir do Critério EB1 (A) e dos Escores do Critério EB1 (B)

Análise dos percentis corrobora o que foi acima apresentado. Foram calculados os escores correspondentes aos percentis 5, 25, 50 e 75 (com base nos escores obtidos na amostra deste estudo) para ambos os critérios em função do total de acertos e por série. Começaremos a análise pelos percentis P25 e P75, pois somente estes foram informados na amostra de padronização do TDE. Conforme se pode observar na Tabela 2, pelo EB1, já na 1ª série foi necessário que a criança atingisse um escore de 51 (ou seja, acertasse 73%

do teste) para que ela fosse enquadrada no percentil 25 (o mesmo escore foi necessário para se atingir o percentil 25 na 3ª série pelo critério EB2). Na quarta série foi necessário que a criança acertasse 90% das palavras do teste (63 acertos) para que fosse colocada no mesmo percentil. Isso torna evidente a necessidade de itens mais difíceis, que possam discriminar os leitores mais competentes entre si. Demonstra também a pouca capacidade do teste em avaliar as diferenças individuais dos menos capazes de uma forma geral (pois tanto o

⁶ A importância desse fato fica evidente ao verificarmos que a escolaridade máxima dos participantes desse estudo é a 4ª série, ao contrário da amostra de padronização, que constitui a 6ª série.

participante que não acertou nenhuma palavra quanto o que acertou 89% das palavras estão na mesma posição relativa – considerando-se o percentil 25, que é o mínimo da amostra de padronização).

Tabela 2: Escore Bruto Mínimo Necessário para Atingir os Percentis 25, 50 e 75 Para os Critérios EB1 e EB2 na Amostra do Presente Estudo

Percentil/ Série	P5		P25		P50		P75	
	EB1	EB2	EB1	EB2	EB1	EB2	EB1	EB2
1ª	37	21	51	36	59	45	63	54
2ª	47	29	56	43	62	53	66	59
3ª	50	44	60	51	63	56	68	64
4ª	56	47	63	58	68	64	69	66
Total	47	34	58	46	63	56	67	63

Nota: P = percentil.

Considerando-se todas as séries juntas, foi necessário acertar 83% dos itens (ou seja, 58) para se atingir o percentil 25, e acertar 95% dos itens (67 itens) para atingir o percentil 75, em relação ao critério EB1. Para os mesmos valores de percentis, o percentual de acerto para o EB2 ficou entre 65,7% (46 itens) e 90% (63 itens). Nota-se, portanto, uma maior distância da proporção de acertos entre os extremos de percentis pelo critério EB2 em relação ao EB1, o que demonstra seu maior poder de informar sobre as diferenças individuais no teste (no caso, as diferenças entre os níveis “inferior” e “superior”). Não obstante, mesmo utilizando-se o novo critério, percebe-se que o problema da distribuição não se resolve completamente, sendo necessário adicionar ao teste itens mais difíceis, para diminuir a proporção de acerto em todos os níveis (superior, médio e inferior).

Outra hipótese seria que o teste apresenta um número excessivo de itens fáceis, o que permite, mesmo a crianças

muito novas ou leitores pouco eficientes, acertarem um grande número de palavras. Isso fica evidente quando avaliamos o percentil 5 nessa amostra. Considerando-se todas as séries juntas, foi necessário, pelo EB1, que o participante acertasse mais da metade do teste (47 acertos ou 67%) para que pertencesse ao referido percentil. O excesso de itens fáceis no teste dificulta a produção de variabilidade entre aqueles indivíduos que são pouco capazes; ou seja, em nossa amostra, o teste não discriminou os participantes que apresentaram um escore de zero dos que conseguiram ler até 47 palavras, embora esses leitores estejam em níveis de desenvolvimento nitidamente diferentes. Pelo EB2, o participante teve que acertar 34 palavras (quase 50% do teste) para pertencer ao percentil 5, o que também é um valor muito elevado.

O impacto do critério de acerto na classificação descritiva dos participantes⁷, de acordo com as normas dadas pela amostra de padronização, pode ser mais bem visualizado na distribuição apresentada pela Figura 2. À primeira vista, pode parecer que o critério adotado em EB2 tornou o teste muito exigente, pois a proporção da classificação “inferior” foi nitidamente maior neste critério, além de ter diminuído significativamente a proporção de indivíduos na classificação “superior”; mas uma análise mais detalhada mostrará que, apesar da forma da distribuição de EB2, este critério torna o teste mais discriminativo. Foi feita uma análise na qual consideramos apenas os indivíduos que obtiveram a classificação “inferior” em EB2, o que totalizou 179 casos. Considerando apenas esses casos, obtivemos a frequência das classificações descritivas que esses participantes apresentaram no critério EB1, conforme se observa na Tabela 3.

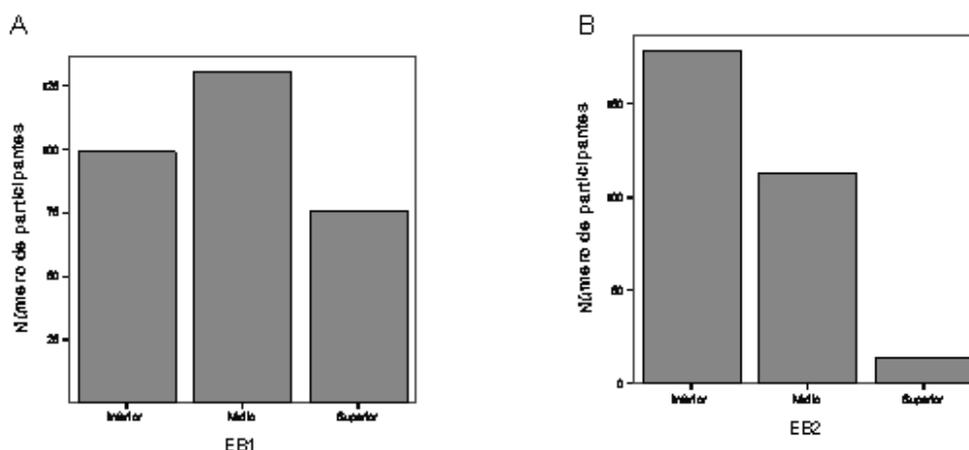


Figura 2. Distribuição da Classificação Descritiva de Acordo com as Normas da Amostra de Padronização para o EB1 (A) e para o EB2 (B)⁷

⁷ Para a primeira série, consideramos sob a classificação de desempenho “Médio” tanto os participantes enquadrados na classificação “Médio Inferior” quanto na classificação “Médio superior”. Tal procedimento foi efetuado para facilitar as análises estatísticas, em virtude de tais classificações não existirem para as demais séries na amostra de padronização

Tabela 3: Frequência da Classificação Descritiva em EB1 Considerando-se Apenas os Participantes Avaliados como “Inferiores” a Partir do Critério Adotado em EB2.

Classificação	Frequência	Porcentagem
Inferior	99	55,3
Médio	63	35,2
Superior	17	9,5

Podemos observar que há uma concordância em torno de 55% da classificação “inferior” entre ambos os critérios adotados. No entanto, em grande proporção, os indivíduos considerados inferiores pelo critério EB2 foram classificados como “médios” pelo critério EB1 (35%) e até mesmo como superiores (9,5%). Isso demonstra que alguns participantes considerados normais pelo critério de Stein (1994) podem ter, na realidade, suas dificuldades camufladas por meio do critério de acerto utilizado.

Outro fato que corrobora a hipótese de que o critério aqui adotado não aumenta demasiadamente o nível de exigência do teste é a distribuição da classificação descritiva de EB2 a partir dos percentis 25 e 75 obtidos na amostra deste estudo, ao invés do estudo de padronização (ver Tabela 2). Com esses percentis obteremos uma distribuição semelhante ao que se demonstra na Figura 2A, conforme se verifica na Figura 3.

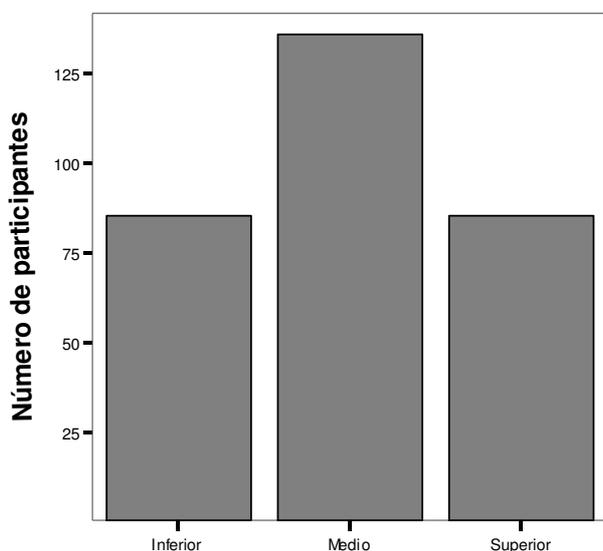


Figura 3. Distribuição das Classificações Descritivas de EB2 (chamada aqui de EB2a), de acordo com os Percentis 25 (inferior) e 75 (superior) Obtidos a Partir dos Resultados da Amostra do Estudo. Note que a Classificação Média é Obtida pelos Valores de Percentis entre 25 e 75.

DISCUSSÃO

Este trabalho demonstrou que o critério de acerto adotado nesta pesquisa para a correção do subteste de leitura TDE repercutiu em diferenças qualitativas e quantitativas nos escores de teste da amostra investigada. Do ponto de vista quantitativo, observou-se que as médias de acerto obtidas por meio do critério do manual (EB1) foram significativamente maiores do que aquelas obtidas por meio do critério adotado nesta pesquisa (EB2), indicando que a adoção de um critério ou de outro provoca diferentes impactos no escore dos indivíduos. Do ponto de vista qualitativo, observou-se que o novo critério proporcionou uma maior variabilidade nos escores do teste, alterando levemente a forma da distribuição, que, de marcadamente assimétrica à direita com o critério de Stein (1994), passou a apresentar uma maior dispersão em torno da média.

Não obstante, o uso do novo critério não impediu a existência do efeito de teto, tampouco proporcionou uma distribuição normal dos escores. A hipótese aqui lançada é que o subteste de leitura do TDE apresenta muitos itens fáceis e poucos itens com um nível maior de dificuldade, os quais seriam capazes de discriminar as habilidades das crianças mais experientes e dos leitores mais capazes. A existência de um excesso de itens fáceis no teste ficou evidente por meio da análise dos percentis obtidos a partir da amostra do presente estudo. Pelo critério EB1, um número muito grande de palavras teve que ser corretamente lido a fim de que o leitor fosse posicionado no percentil 25 (apresentasse um desempenho inferior à média), variando de mais de 70% de acertos na primeira série a 90% na quarta série. Esse fato acarreta sérias repercussões. Uma delas é a possibilidade de que o excesso de itens fáceis provoque um desperdício de recursos no processo de avaliação (como de tempo, ou da colaboração do probando em participar). Desta forma, uma margem de 30% de itens fáceis seria talvez mais apropriada para diferenciar os leitores menos capazes dos mais eficientes. Tal índice foi adotado, por exemplo, por Martin e Pratt (2001) na análise de itens de um teste de leitura de não-palavras isoladas para a língua inglesa. Outro ponto reflete sobre o poder de discriminação do teste. Como o participante deve apresentar um escore muito elevado (em relação à quantidade total de itens existentes) para que seja posicionado ainda na classificação inferior, a leitura correta de poucos itens a mais o leva da classificação média para a superior, o que nos faz pensar se, de fato, o corpo de itens restantes é suficiente para discriminar as habilidades dos leitores médios das dos superiores.

Apesar de não ter eliminado o efeito de teto, o critério EB2 diminuiu esse efeito ao se observar uma menor proporção de indivíduos que acertaram o teste todo ou quase todo, indicando uma maior variabilidade na distribuição dos escores. Essa maior variabilidade também ficou evidente pela exigência de uma menor proporção de acertos para que os indivíduos atingissem o percentil 25 e pela existência de um maior intervalo de escores entre os extremos de percentis, o que permitiu uma maior distância entre os níveis de acertos dos participantes considerados superiores e os daqueles com níveis inferiores na habilidade investigada.

Outro ponto que merece destaque é a proporção de indivíduos que, sendo considerados normais pelo critério de Stein (1994) (obtem a classificação “média” ou “superior”), recaem na faixa inferior de acordo com o critério aqui adotado. A concordância entre as classificações ficou em torno de 55%, mas existe uma grande proporção de indivíduos cujas dificuldades podem ter sido camufladas em função do critério de acerto estipulado pelo manual. Uma possível explicação para essa divergência seria que o novo critério teria aumentado demasiadamente o nível de dificuldade do teste. Esta hipótese não se confirmou, pois a distribuição das classificações descritivas dos escores obtidos por meio do novo critério foi muito semelhante à distribuição dos escores obtidos pelo critério do manual, quando consideramos os percentis gerados a partir da amostra do trabalho ao invés daqueles obtidos pela amostra de padronização.

Por fim, resta responder à questão: por que o subteste de leitura do TDE apresenta um excesso de palavras fáceis? A resposta a esta pergunta pode ser encontrada no nível de regularidade das palavras utilizadas. A lista de palavras do subteste de leitura do TDE foi construída levando em consideração a gradação dos fonemas segundo a relação fonológico-ortográfica das palavras, ou seja, as palavras foram classificadas em relação à correspondência fonema-grafema, que determina a regularidade das palavras para a escrita. Segundo Pinheiro e Rothe-Neves (2001), muitas palavras que são irregulares para a escrita causam pouca ou nenhuma dificuldade para a leitura, pois podem ser corretamente lidas mediante o uso da aplicação de regras de correspondência entre grafemas e fonemas. Essas regras podem ser independentes de contexto (quando existe uma relação unívoca entre a grafia e o som, por exemplo, na palavra *baba*) ou dependentes de contexto (por exemplo, o caso do

<s> e do <r> intervocálicos, que são pronunciados respectivamente como /z/ e /r/). Desta forma, para que a avaliação cognitiva da leitura seja eficaz, deve-se considerar a correspondência grafema-fonema, das palavras, e não a relação fonema-grafema, pois é deste modo que controlamos a regularidade das palavras para a leitura (Parente et al., 1997). Assim, supõe-se que as palavras do subteste de leitura do TDE sejam fáceis em função do excesso de palavras regulares e da pouca proporção de palavras irregulares para a leitura. Isso se confirma ao se constatar que 67% das palavras do teste são regulares para a leitura, sendo que, destas, a grande maioria (73%) apresenta relação unívoca entre grafemas e fonemas e apenas uma minoria (27%) apresenta pronúncia dependente de contexto, o que poderia causar alguma dificuldade para as crianças que ainda não aprenderam regras contextuais. Essas considerações devem, portanto, ser levadas em conta no momento de construir listas de palavras que apresentem um nível de dificuldade adequado para se avaliar a habilidade de reconhecimento de palavras. Além disso, deve-se também considerar o controle tanto da variável frequência de ocorrência de palavras como da variável comprimento de palavras.

A propósito, como nos mostram Pinheiro e Rothe-Neves (2001), uma lista de palavras construída de forma a testar a utilização de estratégias de leitura (fonológica ou lexical) ao longo do desenvolvimento deve, em primeiro lugar, equiparar as palavras em nível de frequência. Assim, tal instrumento deve conter o mesmo número de palavras de alta e baixa frequência de ocorrência e, por sua vez, em cada um desses níveis, igual número de palavras regulares e irregulares. Finalmente, dentro de cada nível de frequência e de regularidade deve haver o mesmo número de palavras curtas e longas. A lista do subteste sob consideração não mostrou essa sutileza, já que, além do predomínio de palavras regulares, houve predomínio de palavras de baixa frequência de ocorrência⁸ e de palavras longas, o que sem dúvida induziu as crianças (principalmente

⁸ De acordo com a contagem de frequência de ocorrência de palavras de Pinheiro (1996). A classificação da frequência das palavras do subteste de leitura do TDE foi feita com base em outro banco de dados (não citado no manual), pois foi publicado antes da publicação do trabalho de Pinheiro. Como a autora não relata a frequência das palavras no manual, utilizamos o banco de Pinheiro para identificar a frequência das palavras apresentadas no subteste.

as de 1ª série) a uma utilização exagerada de leitura silabada.

REFERÊNCIAS

- Capellini, V. L. M., Mendes, E. G., & Salgado, H. M. (2003). Avaliando o rendimento escolar de alunos com deficiência auditiva em escolas regulares. *Revista espaço*, 18-19, dezembro de 2002 a julho 2003.
- 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.
- Ehri, L. C. (1992). Reconceptualizing the development of sight word reading and its relationship to reading. Em: P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading Acquisition* (pp. 35-48). Hillsdale (NJ): Lawrence Erlbaum.
- Guimarães, S. R. K. (2003). Dificuldades no desenvolvimento da lectoescrita: o papel das habilidades metalingüísticas. *Psicologia: teoria e pesquisa*, 19 (1), 33-45.
- Martin, F., & Pratt, C. (2001). *The Martin and Pratt nonword reading test*. Melbourne: Australian Council for Educational Research.
- Parente, M. A. M. P., Silveira, A., & Lecours, A. R. (1997). As palavras do português escrito. Em: A. R. Lecours & M. A. M. Parente (Eds.), *Dislexia: Implicações do Sistema de Escrita do Português* (pp. 41-55). Porto Alegre: Artes Médicas.
- 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 na faixa pré-escolar e séries iniciais do 1º grau*. ABD – Associação Brasileira de Dislexia, São Paulo.
- Pinheiro, A. M. V. & Rothe-Neves, R. (2001). Avaliação cognitiva da leitura: as tarefas de leitura em voz alta e ditado. *Psicologia: Reflexão e Crítica*, 14 (2), 399-408.
- Salles, J. F. (2005). *Habilidades e dificuldades de leitura e escrita em crianças de 2ª série: abordagem neuropsicológica cognitiva*. Tese de doutorado não publicada. Programa de Pós-Graduação em Psicologia do Desenvolvimento, Universidade Federal do Rio Grande do Sul, Porto Alegre.
- Scliar-Cabral, L. (2003) *Princípios do Sistema Alfabético do Português do Brasil*. São Paulo: Contexto.
- SPSS (2006). *Statistical Package for the Social Sciences for Windows (Version 15.0)* [Computer software]. Chicago: Author.
- Stein, L. M. (1994). TDE: *Teste de Desempenho Escolar*: manual para aplicação e interpretação. São Paulo: Casa do Psicólogo.
- Stuart, K. M. & Coltheart, M. (1988). Does reading develop in a sequence of stages? *Cognition*, 30, 139-181.

Recebido em 27/05/2008

Aceito em 04/02/2009

Endereço para correspondência : Patrícia Silva Lúcio. Rua Costa Rica, n. 455, Bairro Nações Unidas, CEP 34590-380, Sabará-MG, Brasil. E-mail: pslucio@gmail.com

