

Propriedades Psicométricas da Prova de Reconhecimento de Palavras

Psychometric Properties of the Word Recognition Test

Fernanda Leopoldina Viana^{*, a}, Iolanda da Silva Ribeiro^a, José Maia^b & Sandra Santos^a

^aUniversidade do Minho, Braga, Portugal & ^bUniversidade do Porto, Porto, Portugal

Resumo

Este estudo tem por objetivo apresentar dados relativos à fidelidade e validade da Prova de Reconhecimento de Palavras (PRP), destinada a alunos dos 4 primeiros anos de escolaridade. A prova é constituída por 40 palavras regulares de 2 e 3 sílabas. Realizaram-se dois estudos. Um primeiro, com 328 alunos, e um segundo com 3.131. Os coeficientes de consistência interna foram elevados nos 4 anos (variando entre 0,96 e 0,98). A diferença na média de desempenho entre anos foi estatisticamente significativa, com valores moderados de magnitude do efeito. Não se encontraram diferenças estatisticamente significativas entre sexos. As correlações com critérios externos foram positivas e estatisticamente significativas. Os dados da AFC são ambíguos, apontando para a necessidade de realização novos estudos.

Palavras-chave: Leitura, reconhecimento de palavras, avaliação psicológica.

Abstract

This study's purpose is to present data regarding the reliability and validity of the Word Recognition Test (WRT) for students of the first 4 years of schooling. The test consists of 40 regular words of 2 and 3 syllables. Two studies were performed. The first included 328 students, and the second one 3,131 students. The coefficients of internal consistency were high in the whole 4 years (ranging between .96 and .98). The difference on performance average was statistically significant between years, with moderate magnitude effect. No statistically significant differences were found between sexes. Correlations with external criteria were positive and statistically significant. Results of the Confirmatory Factor Analysis (CFA) are not conclusive, pointing to the need of further studies.

Keywords: Reading, word recognition, psychological assessment.

A capacidade para identificar e reconhecer as palavras escritas de forma rápida e precisa é determinante para a compreensão da leitura. A automatização deste reconhecimento é produto de um processamento coordenado de informações sobre a pronúncia das palavras, os seus significados e sobre as identidades ortográficas das mesmas (McGuiness, 2005; Morais, 1997; Snowling & Hulme, 2005).

A partir dos anos 70 do século XX surgiram diferentes modelos explicativos do processamento de informação envolvido no reconhecimento de palavras, adotando diferentes conceptualizações sobre o léxico mental e o seu acesso. Numa primeira conceção o léxico mental é entendido como uma espécie de ficheiro, no qual as entradas são estruturadas de acordo com um conjunto bem definido de critérios. As representações lexicais são de natureza estática e o trabalho de reconhecimento de uma palavra é idêntico ao que se efetua na pesquisa de um dicionário. Trata-se de um modelo ativo de acesso ao léxico, dado que a localização de um item lexical exige uma procura ativa

no conjunto de todas as entradas. Numa segunda conceção as palavras são representadas por unidades internas (detetores), suscetíveis de serem ativadas pela informação sensorial proveniente da palavra estímulo. O reconhecimento de uma palavra ocorre quando é atingido um determinado limiar de ativação de uma unidade interna. Contrariamente à anterior, esta conceção não prevê a existência de um procedimento ativo de procura, configurando um modelo passivo (Segui, 1991).

A inclusão de um terceiro critério – o tipo de informação suscetível de interferir no acesso ao léxico – permitiu um segundo tipo de classificação. Os modelos de acesso serão autónomos, se consideram que apenas as informações de nível inferior de tratamento intervêm na determinação do acesso (Becker, Goldinger, & Stone, 2006), ou interativos, se aceitam que possam também ser consideradas informações provenientes de níveis superiores de análise (Bar et al., 2006). Por exemplo, no modelo passivo e iterativo de Morton (1969, 1982) a identificação de uma palavra tem lugar quando o número de índices disponíveis (evidências) é suficiente para que o detetor (logogene) atinja o seu limiar de ativação.

Esta distinção clássica sustentou várias hipóteses de investigação. Uma primeira foi a de considerar que a

* Endereço para correspondência: Departamento de Psicologia da Educação e Educação Especial, Instituto de Educação, Universidade do Minho, Campus de Gualtar, Braga, Portugal 4710-057. E-mail: fviana@ie.uminho.pt

dificuldade de reconhecimento de uma palavra estaria ligada à extensão da classe de candidatas que poderiam ser ativadas face a um determinado estímulo sensorial (Coltheart, Davelaar, Jonasson, & Besner, 1977). De acordo com esta hipótese, o reconhecimento de uma palavra exige distingui-la de todas as outras do léxico e, em particular, das suas vizinhas ortográficas. No conceito de vizinhança ortográfica proposto por estes autores são incluídas todas as outras palavras da língua, com a mesma extensão, que partilham com ela todas as letras, à exceção de uma, nas suas posições sequenciais respetivas, como, por exemplo “bota” e “bola”. Os resultados do estudo destes autores não mostrou a existência de correlações entre o tempo de resposta na leitura de uma palavra e o número das suas vizinhas ortográficas, provavelmente porque apenas levaram em linha de conta a variável extensão dos candidatos.

De acordo com as propostas de Forster (1981) e de Paap, Newsome, McDonald e Schvaneveldt (1982), as vizinhas ortográficas não são examinadas de forma aleatória, mas de acordo com a sua frequência. Segundo outros modelos interativos (e.g. Rumelhart & McClelland, 1982), quanto mais elevada é a frequência de uma palavra, maior é a força inibitória que exerce sobre as suas vizinhas. Assim sendo, a dificuldade de reconhecimento de uma palavra depende da sua frequência relativamente às suas vizinhas ortográficas (Grainger, O’Regan, Jacobs, & Segui, 1989). No entanto, e como mostrou o estudo de Grainger e Segui (1990), embora as palavras sem vizinhas ortográficas mais frequentes sejam lidas mais rapidamente, a análise qualitativa dos erros não foi conclusiva, dado que estes nem sempre correspondiam à escolha de uma palavra considerada vizinha ortográfica.

O efeito da vizinhança ortográfica tem sido bastante estudado na psicologia experimental (Andrews, 1989, 1992, 1997; Grainger et al., 1989; Justi & Pinheiro, 2006; J. Perea, 2008; J. Perea & Rosa, 2000; M. Perea, Carreiras, & Grainger, 2004, entre outros) mas não tem integrado a avaliação da leitura em larga escala.

As tarefas mais utilizadas para analisar o efeito da vizinhança ortográfica são as de identificação perceptual (apresentam-se as palavras durante um curto período de tempo, acompanhadas de uma “máscara” que dificulta a sua leitura e que pode ir sendo gradualmente diminuída), as de decisão lexical (na qual os sujeitos têm de decidir se um determinado conjunto de letras é ou não uma palavra) e as de leitura de palavras em voz alta, geralmente apresentadas na tela de um computador durante um determinado período de tempo (Grainger, 2002). As primeiras têm a desvantagem de induzir estratégias de adivinhação, nomeadamente se o sujeito sabe da existência de tempo limite para a efetuação da prova (Andrews, 1997; Carreiras, Perea, & Grainger, 1997). Quanto à tarefa de leitura em voz alta, a principal limitação tem a ver com o fato de o sujeito poder aceder à forma fonológica das palavras (identifique) sem que ative qualquer representação semântica (compreenda).

Em Portugal (Sim-Sim & Viana, 2007), bem como no Brasil (Lúcio, 2008), existem poucas provas que permitam

avaliar o reconhecimento de palavras. A Prova de Reconhecimento de Palavras (PRP) pretende contribuir para colmatar esta lacuna. No seu desenvolvimento procurou-se conciliar uma avaliação de reconhecimento de palavras de acordo com a conceptualização teórica de léxico mental interativo e a hipótese da ativação de múltiplos candidatos lexicais ao processo de ativação com o uso de um tipo de tarefa que pudesse contornar os constrangimentos atrás referidos. Procurou-se ainda responder à necessidade de criar uma prova de *screening*, destinada a ser usada em avaliações em larga escala e em contexto escolar, não exigindo o recurso a equipamento e *software* específicos, sendo passível de aplicação coletiva.

A opção por construir uma tarefa de *screening* justificou-se pela inexistência, em Portugal, de provas validadas para a avaliação de leitura de palavras isoladas (Sim-Sim & Viana, 2007) o que, aliado à inexistência de marcos de referência que permitam decidir se o desempenho observado pode, ou não, ser considerado o expectável num ano de escolaridade constitui uma enorme limitação no processo de diagnóstico e de intervenção.

Atendendo a estes objetivos, a PRP usa um formato original de tarefa. Em cada item é apresentada uma imagem, seguida de 4 palavras, devendo os sujeitos assinalar a palavra que nomeia a imagem. Em 5 itens são apresentadas 3 vizinhas da palavra-alvo (ex: badana, baiana, bacana, *banana*), em 16 itens são apresentadas 2 vizinhas da palavra-alvo (ex: capelo, *camelo*, cabelo, caneco), em 13 itens é apresentada 1 vizinha da palavra-alvo (ex: morando, molengo, nadando, *morango*) e em 6 itens apenas são apresentadas palavras com proximidade ortográfica (ex: censura, cegonha, tesoura, *cenoura*). A palavra alvo é sempre a mais frequente. Não exige leitura em voz alta, tendo o sujeito apenas de marcar a palavra correspondente à imagem fornecida. Dado tratar-se de uma prova de *screening*, a PRP integra apenas palavras regulares.

A PRP foi construída procurando proporcionar quer a investigadores quer a professores ou psicólogos uma medida fiel e válida que permita avaliar o desempenho dos alunos dos primeiros quatro anos de escolaridade ao nível do reconhecimento de palavras regulares.

A presente versão da PRP é composta por 40 itens, que avalia a leitura silenciosa de palavras, contemplando em simultâneo a precisão e a velocidade de leitura. Em 20 itens são usadas palavras com 2 sílabas e, nos outros 20, palavras com 3 sílabas. Relativamente às características psicolinguísticas dos estímulos, foi controlada a frequência, a regularidade, a extensão e a vizinhança ortográfica. Todas as palavras alvo são de frequência superior às vizinhas ortográficas. É atribuído um ponto por cada item corretamente respondido. No 1.º e 2.º anos de escolaridade o tempo máximo para a realização da prova é de 4 minutos; no 3.º e 4.º anos é de 2 minutos.

Num estudo piloto inicial (Viana & Ribeiro, 2006), efetuado com 272 crianças do 1.º Ciclo do Ensino Básico, distribuídas pelas quatro séries iniciais, e integrando 77 itens, foi avaliada a iconicidade das imagens relativas às

palavras-alvo, tendo sido eliminados os itens cujas imagens não foram nomeadas com recurso às palavras-alvo. Após esta eliminação, a versão experimental da prova passou a ser constituída por 40 itens. A localização da palavra-alvo foi aleatorizada nas 4 alternativas (5 vezes em primeiro lugar, 11 vezes em segundo e quarto lugares e 13 vezes em terceiro lugar). Nesta versão, a correlação corrigida item-total variou entre 0,40 e 0,88; o valor médio das correlações situou-se em 0,66. O valor médio do índice de dificuldade foi de 0,73, com uma variação entre 0,46 e 0,95. A média das intercorrelações entre os itens da prova é de 0,67 (Viana & Ribeiro, 2007). Neste estudo procedeu-se igualmente à análise dos erros, tendo-se verificado que as escolhas incorretas não apresentavam correlações significativas do ponto de vista estatístico nem com a frequência relativa das vizinhas ortográficas, nem com a posição da palavra-alvo. Embora a PRP use um tipo de tarefa diferente das usadas nos estudos revistos por Andrews (1997), os resultados encontrados corroboram as suas conclusões que apontam no sentido de os efeitos do número de vizinhas ortográfica e da sua frequência só serem encontrados, de forma consistente, com palavras de baixa frequência, o que não é o caso da seleção efetuada para as palavras que integram esta prova.

A revisão de estudos centrados no reconhecimento de palavras por crianças falantes do português, efetuada por Lúcio e Pinheiro (2011), permite-nos concluir que os mesmos são conduzidos numa ótica experimental, procurando analisar como as características psicolinguísticas das palavras influenciam a sua leitura e de que modo estas diferenças variam em função da escolarização. Todavia, verifica-se também que os mesmos são desenvolvidos sem terem como objetivo a construção de provas que permitam comparar o desempenho dos alunos no reconhecimento de palavras com o esperado para o seu ano de escolaridade ou com um critério de mestria, não procedendo à análise das características psicométricas das diferentes listas de palavras usadas em termos de fidelidade e de validade. Nas provas publicadas em português que visam avaliar o reconhecimento de palavras (Capovilla, Varanda, & Capovilla, 2006; Stein, 2008) não é colocada qualquer hipótese sobre a estrutura fatorial das mesmas. Da revisão efetuada, somente Lúcio (2008) discute este tópico, tendo interpretado os valores elevados do coeficiente de consistência interna como indicando unidimensionalidade. Porém, esta hipótese não foi testada com recurso a análise fatorial.

Este estudo tem como objetivo analisar as características psicométricas da PRP, em particular a fidelidade e validade dos resultados. Para o efeito efetuaram-se dois estudos complementares, os quais visam analisar a consistência interna, a validade de constructo e a validade concorrente.

Estudo 1

Como referido, o reconhecimento de palavras tem sido abordado mais numa vertente experimental do que

psicométrica. Este facto explicará que na revisão efetuada se encontrem poucas referências ao que poderá ser expectável em termos de estrutura fatorial. Atendendo às características dos itens que integram a PRP, a hipótese a testar é a da unidimensionalidade medida. O teste desta hipótese é o primeiro objetivo do Estudo 1. O segundo objetivo deste estudo prende-se com o cálculo da consistência interna da PRP.

Método

Participantes

Participaram neste estudo 328 alunos de ambos os sexos (59,1% sexo masculino e 40,9% do sexo feminino) do 1.º Ciclo do Ensino Básico. Todos os alunos frequentavam escolas públicas – rurais (21,6%) e urbanas (78,4%) – do norte de Portugal, tinham o português como língua materna e nenhum apresentava necessidades educativas especiais de carácter permanente. Em termos de idades, as médias situavam-se em 6,8 ($DP = 0,58$), 7,5 ($DP = 0,60$), 8,5 ($DP = 0,51$) e 9,6 anos ($DP = 0,52$) respetivamente para o 1.º, 2.º, 3.º e 4.º ano. Em Portugal as crianças ingressam na escola no ano em que completam os 6 anos.

Medidas

Neste estudo foi usada a PRP, atrás descrita.

Procedimentos

Foram consideradas as orientações éticas preconizadas pela *American Psychological Association*. A aplicação das provas foi precedida da obtenção das autorizações necessárias junto das Direções das escolas e dos Pais, após informação sobre os objetivos do estudo. A participação dos alunos era voluntária, tendo sido garantida a confidencialidade dos resultados. A prova foi aplicada de forma coletiva, sem limite de tempo, durante o tempo letivo, por psicólogos especialmente treinados.

Análise Estatística

Analisou-se a dimensionalidade da PRP efetuando uma análise fatorial confirmatória com o programa Mplus (L. K. Muthén & Muthén, 1998-2010). Na avaliação do ajustamento global do modelo usaram-se as seguintes estatísticas: qui-quadrado (χ^2), o *Root Mean Square Error of Aproximation* (RMSEA), o *Weighted Root Mean Square Residual* (WRMR), o *Comparative Fit Index* (CFI) e o *Tucker-Lewis Index* (TLI; Hu & Bentler, 1999). Dado que as respostas aos itens da PRP são de natureza dicotómica (certo/errado) recorreu-se aos procedimentos implementados no programa Mplus. O teste de χ^2 é um teste formal à significância da função de discrepância entre a matriz de covariância populacional e a postulada pelo modelo. Quanto menor for o seu valor relativamente ao número de graus de liberdade melhor é o ajustamento. O RMSEA é um índice de discrepância populacional que penaliza modelos com excesso de parâmetros. Valores superiores

a 0,08 sugerem um modelo com um ajustamento pobre. O WRMR baseia-se nas diferenças médias entre a matriz de correlações observada e a matriz de correlações prevista pelo modelo. É um índice adequado para dados binários (Cook, Kallen, & Amtmann, 2009), introduzido por L. K. Muthén e Muthén (1998-2010). Não sendo estandardizado pode assumir um valor situado para além do intervalo 0,0-1,0. Os valores inferiores à unidade são considerados bons indicadores de ajustamento do modelo (Yu, 2002). O *Comparative Fit Index* (CFI) e o *Tucker-Lewis Index* (TLI) são, respetivamente, índices de ajustamento incremental e de parcimónia do modelo relativamente a um modelo nulo. São considerados aceitáveis valores superiores a 0,90 (Bentler & Bonett, 1980; Hu & Bentler, 1999; Kline, 2005). O coeficiente *alpha* de Cronbach foi calculado para testar a consistência interna da PRP. Usou-se a versão 18 do SPSS.

Resultados

Uma vez que a escala obriga a uma resposta binária em todos os itens recorreu-se a um estimador robusto proposto por B. O. Muthén, du Toit e Spisic (1997) e implementado na versão mais recente do *software* Mplus. Neste sentido, o valor de χ^2 foi obtido do estimador robusto WLSMV, e o mesmo aconteceu com os outros índices de ajustamento global. O valor do qui-quadrado é estatisticamente significativo ($\chi^2=335,248$; $gl=51$; $p<0,001$). Num sentido estrito este valor indica que o ajuste do modelo não é satisfatório. Contudo, o valor deste teste per se não é suficiente para apreciar o ajuste do modelo aos dados empíricos, dado que o mesmo é influenciado pela dimensão amostral. Em amostras com números de casos superiores a 200, como é o caso do presente estudo, o teste de qui-quadrado é quase sempre estatisticamente significativo (Motl & DiStefano, 2002).

A fim de reduzir o efeito do tamanho da amostra no teste de qui-quadrado, Kline (2005) sugere o cálculo de um quociente entre o valor do teste e os graus de liberdade. O valor deste quociente é, neste estudo, igual a 6,51. Embora não haja ainda consenso sobre o valor mínimo desta razão, com autores a sugerir como aceitáveis valores inferiores a 2,00 e outros a 3,0 (Blunch, 2008; Kline, 2005), o valor da razão obtido é superior aos mínimos avançados.

O RMSEA é igual a 0,13, superior ao valor crítico aceite de ajuste do modelo (Brown & Cudeck, 1993). Igualmente o valor do WRWR (1,650) é superior ao sugerido por L. K. Muthén e Muthén (1998-2010). Por sua vez os índices de ajustamento global obtidos são elevados (CFI=0,979; TLI=0,989), ultrapassando os valores de referência (Hu & Bentler, 1999). Na Tabela 1 apresentam-se as saturações de cada um dos itens, as quais são todas significativamente superiores a zero.

O valor do coeficiente de consistência interna é de 0,98, situando-se acima do valor de referência (0,70) proposto por Nunnally (1978).

Discussão

Os resultados da Análise Fatorial Confirmatória apresentam-se contraditórios. Por um lado o valor do teste de qui-quadrado, do RMSEA, bem como do WRWR sugerem problemas no ajustamento global do modelo. Por outro lado, considerando outros indicadores de ajustamento (CFI e TLI) os resultados são, por sua vez, superiores aos pontos de corte usualmente referidos na literatura. Os valores observados poderão estar associados a aspectos idiossincráticos dos itens que constituem esta prova. Há muitos itens em que as probabilidades de dar respostas certas são muito elevadas. No cálculo do coeficiente de correlação tetracórico (usado neste caso face à natureza binária das respostas) utilizam-se, necessariamente, as frequências conjuntas de itens. Dada a elevada probabilidade de acertos nos itens esperava-se a ocorrência de correlações muito próximas da unidade.

O valor de consistência interna encontrado é similar ao encontrado por Lúcio (2008) junto de uma amostra com alunos de várias séries, que foi de 0,99.

Uma das limitações do presente estudo prende-se com a estimação da fidelidade com um grupo que inclui alunos de vários anos de escolaridade. Nem sempre o mesmo teste apresenta valores de fiabilidade adequados quando são avaliados sujeitos que diferem em termos de variáveis como a idade ou o ano de escolaridade (Feldt & Kim, 2006), pelo que é necessário calcular a consistência para os vários grupos. Este aspecto foi considerado no segundo estudo.

Estudo 2

O objetivo deste estudo foi o de analisar a fidelidade (consistência interna) dos resultados da prova junto de alunos de diferentes anos de escolaridade e analisar a sua validade concorrente. Especificamente procedeu-se a estudos de correlação com provas que avaliam o mesmo constructo, tendo-se igualmente testado a existência de diferenças no desempenho em função do sexo e do ano de escolaridade.

Método

Participantes

Neste estudo foram seguidos os mesmos procedimentos amostrais do Estudo 1, mas alargou-se a recolha de dados às zonas centro e sul de Portugal. A amostra foi constituída por 3131 alunos, cuja distribuição, em função das variáveis idade, sexo e zona geográfica, pode ser observada na Tabela 2. Nos 4 anos de escolaridade as médias de idades foram, respetivamente de 6,6 ($DP=0,47$; amplitude: 6,6 – 7,6), 7,5 ($DP=0,49$; amplitude: 7,0 – 9,1), 8,6 ($DP=0,52$; amplitude: 8,0 – 10,1) e 9,6 anos ($DP=0,49$; amplitude: 9,0 – 11,1).

Tabela 1
Saturação dos Itens no Fator

Item	Sat.	EP	Sat./ EP	<i>p</i>	Item	Sat.	EP	Sat./ EP	<i>p</i>
1	0,642	0,087	7,422	0,001	21	0,941	0,019	48,91	0,001
2	0,543	0,073	7,40	0,001	22	0,958	0,016	59,38	0,001
3	0,756	0,076	9,921	0,001	23	0,963	0,015	65,62	0,001
4	0,574	0,072	7,92	0,001	24	0,968	0,012	79,907	0,001
5	0,568	0,074	7,62	0,001	25	0,991	0,008	124,39	0,001
6	0,164	0,074	2,22	0,05	26	0,944	0,014	67,70	0,001
7	0,756	0,083	90,09	0,001	27	0,974	0,011	92,09	0,001
8	0,711	0,084	8,50	0,001	28	0,966	0,012	82,36	0,001
9	0,773	0,079	9,73	0,001	29	0,960	0,012	79,90	0,001
10	0,515	0,077	6,712	0,001	30	0,971	0,010	100,74	0,001
11	0,838	0,066	12,67	0,001	31	0,995	0,005	192,78	0,001
12	0,468	0,074	6,35	0,001	32	0,980	0,007	137,55	0,001
13	0,831	0,059	13,98	0,001	33	0,984	0,006	174,16	0,001
14	0,821	0,055	14,91	0,001	34	0,978	0,008	121,11	0,001
15	0,909	0,050	18,22	0,001	35	0,969	0,007	134,55	0,001
16	0,582	0,058	10,07	0,001	36	0,964	0,008	124,99	0,001
17	0,672	0,059	11,34	0,001	37	0,972	0,009	111,98	0,001
18	0,851	0,040	21,09	0,001	38	0,972	0,009	104,80	0,001
19	0,920	0,026	35,27	0,001	39	0,814	0,069	11,87	0,001
20	0,921	0,024	38,45	0,001	40	0,835	0,038	21,78	0,001

Nota. Sat. – Saturação; EP – Erro Padrão.

Tabela 2
Descrição dos Participantes

		1.º Ano		2.º Ano		3.º Ano		4.º Ano	
		<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Sexo	Masc.	210	52,2	471	50,5	488	51,7	436	51,1
	Fem.	192	47,8	461	49,5	456	48,3	418	48,9
Zona	Norte	206	51,2	361	38,7	389	41,3	356	41,7
	Centro	72	17,9	344	36,9	304	32,2	276	32,3
	Sul	124	30,8	227	24,4	250	26,5	222	26,0
Meio	Urbano	332	82,6	789	84,7	821	87,1	722	84,5
	Rural	70	17,4	143	15,3	122	12,9	132	15,5

Medidas

Neste segundo estudo, para além da PRP foram usados os seguintes instrumentos: TIL – Teste de Idade de Leitura; Avaliação do desempenho em leitura efetuada pelos

professores. O TIL (Sucena & Castro, 2008) é uma prova constituída por 36 frases as quais devem ser completadas com a escolha de uma palavra a selecionar de entre 5 alternativas fornecidas. É destinada a alunos do 2.º ao 5.º

ano de escolaridade. A aplicação pode ser individual ou coletiva, com um tempo limite de 5 minutos. A avaliação da qualidade da leitura foi efetuada pedindo aos professores o preenchimento de uma escala tipo Likert de 5 pontos (1-não lê, 5- lê muito bem).

Procedimentos

No que respeita à recolha de dados, os procedimentos foram idênticos aos do Estudo 1. A sequência de aplicação das provas foi a mesma para todos os sujeitos.

Análises Estatísticas

Calcularam-se as estatísticas descritivas relativas aos resultados da PRP em função do ano e do sexo. O estudo da consistência interna foi efetuado recorrendo ao método da bipartição (divisão em itens pares e ímpares). A opção por este procedimento está relacionada com as condições de administração da prova. A realização com limites de tempo leva a que muitos alunos (os que apresentam maiores dificuldades no reconhecimento automático de palavras) não concluem a prova. O método de divisão adotado per-

mite distribuir de modo aleatório os dados omissos pelas duas metades.

A validade concorrente da prova foi examinada através do cálculo do coeficiente de correlação entre os resultados da PRP e os resultados no TIL e com a avaliação dos professores sobre o desempenho em leitura dos alunos. Comparou-se a diferença de médias entre o 1.º e o 2.º anos e entre o 3.º e o 4.º anos através de um teste *t* para amostras independentes, tendo-se calculado a magnitude do efeito com base na fórmula de Cohen (1988). Dado que o tempo concedido para a realização da PRP não é constante (4 minutos para 1.º e 2.º anos e 2 minutos para 3.º e 4.º anos) a comparação é efetuada entre os resultados do 1.º com o 2.º ano e os de 3.º com os de 4.º.

Resultados

Na Tabela 3 apresentam-se os valores das estatísticas descritivas (*M* – Média, *DP* – Desvio Padrão, Amp. – Amplitude de valores mínimos e máximos) em função do ano de escolaridade e do sexo. Não se registam, nos 4 anos de escolaridade, diferenças estatisticamente significativas nas médias de desempenho em função do sexo.

Tabela 3

Média, Desvio Padrão e Amplitude de Resultados na PRP em Função do Ano de Escolaridade e do Sexo

Ano	Sexo	<i>n</i>	<i>M</i>	<i>DP</i>	Amp.	<i>t</i>	<i>gl</i>	<i>p</i>
1	Masc.	210	12,6	8,28	0-40	0,668	400	n.s.
	Fem.	192	13,1	8,19	0-40			
2	Masc.	471	29,9	9,01	3-40	0,946	930	n.s.
	Fem.	461	29,4	9,14	1-40			
3	Masc.	488	27,6	9,11	4-40	-0,24	941	n.s.
	Fem.	455	27,7	9,30	4-40			
4	Masc.	436	32,7	7,41	7-40	0,946	852	n.s.
	Fem.	418	32,5	7,83	10-40			

Nota. *M* – Média; *DP* – Desvio Padrão; Amp. – Amplitude de resultados; n.s. – não significativo.

Os valores do coeficiente de consistência interna, calculados separadamente para cada ano, foram os seguintes: 0,96 (1.º ano); 0,98 (2.º ano); 0,97 (3.º ano) e 0,98 (4.º ano).

Calculou-se a diferença de médias nos resultados da PRP entre os alunos do 1.º e 2.º anos e entre os de 3.º e 4.º anos. Na Tabela 4 apresentam-se os resultados do *t-test* para amostras independentes e o valor da magnitude do efeito (ME). As diferenças entre o 1.º e o 2.º ano, bem como as diferenças entre o 3.º e o 4.º ano são estatisticamente significativas. O valor da magnitude do efeito é elevado para a diferença entre os dois primeiros anos, mas mode-

rado para a diferença de resultados entre o 3.º e o 4.º anos.

O TIL só pode ser aplicado a partir do 2.º ano, razão pela qual não foi calculado este coeficiente para o 1.º ano. Os coeficientes de correlação entre os resultados na PRP e no TIL são estatisticamente significativos em cada um dos anos de escolaridade (2.º ano: *r* = 0,63, *p* < 0,001; 3.º ano: *r* = 0,34, *p* < 0,001; 4.º ano: *r* = 0,38, *p* < 0,001). Os coeficientes de correlação entre os resultados obtidos na PRP e a avaliação do desempenho em leitura realizada pelos professores foram estatisticamente significativos em todos os anos de escolaridade (1.º ano: 0.69, *p* < 0,001;

Tabela 4
 Comparação dos Resultados na PRP em Função do Ano

Ano	<i>n</i>	<i>M</i>	<i>DP</i>	<i>t</i>	<i>gl</i>	<i>p</i>	ME
1	402	12,8	8,22	31,98	1332	0,001	1,91
2	932	29,7	9,07				
3	948	27,6	9,19	12,1	1795	0,001	0,58
4	854	32,5	7,62				

Nota. *n* – número de casos; *M* – Média; *DP* – Desvio Padrão.

2.º ano: 0,77, $p < 0,001$; 3.º ano: 0,63, $p < 0,001$; 4.º ano: 0,67, $p < 0,001$).

Discussão

Os coeficientes de consistência interna são elevados e similares nos 4 anos de escolaridade indicando que a prova permite a medida do reconhecimento de palavras de modo igualmente fidedigno nos anos a que se destina. Valores também muito elevados foram obtidos noutras provas (Gilmore, Croft, & Reid, 1981; Lúcio, 2008). Embora as tarefas propostas sejam diferentes (leitura, em voz alta, de palavras isoladas) os valores dos coeficientes de consistência interna apresentados por Lúcio variaram entre 0,97 na 3ª série e 0,99 na 1ª série. Por sua vez, no *Burt Word Reading Test* (Gilmore et al., 1981) variaram entre 0,96 e 0,97.

As diferenças entre as médias são estatisticamente significativas, sendo os valores de magnitude do efeito elevados entre o 1.º e 2.º anos e moderadas entre o 3.º e 4.º anos, sugerindo que a prova permite a discriminação entre o desempenho dos alunos de vários anos no reconhecimento de palavras.

No Teste de Competência de Leitura de Palavras e Pseudo-palavras (Capovilla et al., 2006) registaram-se efeitos da série escolar na pontuação geral e nos 7 sub-testes, com diferenças estatisticamente significativas da 1ª para a 2ª série e da 2ª para a 3ª.

As correlações verificadas relativamente aos dois critérios externos usados foram positivas e estatisticamente significativas nos quatro anos de escolaridade. As correlações encontradas entre a fluência (avaliada através da leitura em voz alta de um texto corrido – *full text*) e compreensão em leitura oscilam, noutros estudos, entre 0,42 e 0,68 (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Mather, Hammill, Allen, & Roberts, 2004; Yovanoff, Duesbery, Alonzo, & Tindal, 2005). Na pesquisa de Jenkins, Fuchs, van den Broek, Espin e Deno (2003) verificou-se uma correlação de 0,83 entre medidas de fluência (texto corrido) e de compreensão, mas de apenas 0,53 quando a medida de fluência era constituída por uma lista de palavras. No presente estudo os coeficientes de correlação entre a PRP e o TIL situaram-se abaixo destes valores no 3.º e no 4.º anos. Estas diferenças

de resultados podem estar associadas a aspectos distintos das duas provas. A PRP avalia velocidade e fluência de leitura, enquanto o TIL é um teste compósito de compreensão (necessária para a seleção da palavra adequada para completar as frases dadas) e velocidade de leitura, já que é uma prova com tempo limite (5 minutos). Ora a fluência e a velocidade de leitura apresentam geralmente correlações mais elevadas com as medidas de compreensão nos anos iniciais de escolaridade, perdendo o seu poder explicativo à medida que a escolaridade avança, resultante da automatização dos processos de descodificação. À medida que a escolaridade avança, processos mais elevados, como os de inferência e auto-monitoramento, tornam-se cruciais. As correlações mais elevadas entre os resultados da PRP com a avaliação dos professores do que com os do TIL refletirá as diferenças nos dois critérios externos. Aos professores foi explicitamente solicitado que avaliassem a proficiência em leitura no sentido estrito de velocidade e precisão. O TIL, por sua vez mede, em simultâneo, compreensão, precisão e velocidade. O valor das correlações encontradas entre os resultados da PRP e a avaliação dos professores é congruente com o de outros estudos que analisaram a validade de medidas baseadas no julgamento de professores (Hoge & Caladarci, 1989; White, 2004).

Os coeficientes de correlação obtidos entre os resultados na PRP e a avaliação de desempenho em leitura efetuada pelos professores é, no 1.º ano de escolaridade, muito superior ao de um outro estudo efetuado em língua portuguesa por Nikaedo, Kuriyama e Macedo (2007), que se situou em 0,32. Estes coeficientes são também superiores aos encontrados no estudo de Milanowski, Kimball e White (2004) e de White (2004) que se situaram entre 0,21 e 0,50. Esta superioridade pode estar relacionada com os instrumentos usados e com o número de casos analisados. No estudo de Nikaedo et al. (2007) os professores classificam os alunos em função da fase em que se encontram em termos de psicogénese da escrita, com uma amostra de 52 crianças. Nos outros dois estudos a avaliação dos professores incide sobre o desempenho escolar mais geral e não especificamente sobre o reconhecimento de palavras.

Como era esperado, não se registaram diferenças em função do sexo nos 4 anos a que se destina a prova,

resultados similares aos encontrados noutros estudos (Johnson, 1973; Walla, Hufnagl, Lindinger, Deecke, & Lang, 2001).

Conclusão

Esta investigação, organizada em dois estudos, desenvolvidos com amostras distintas, facultou informação sobre o desenvolvimento da PRP – Prova de Reconhecimento de Palavras e as suas propriedades psicométricas – fidelidade e validade.

Nos dois estudos efetuados, os valores de consistência interna foram elevados e similares nos quatro anos de escolaridade considerados. Este aspecto é de realçar, uma vez que é conhecido que uma mesma prova quando aplicada em anos diferentes pode apresentar valores de consistência interna diferentes indicando que a prova não é igualmente fidedigna nos vários anos (Feldt & Kim, 2006).

A unidimensionalidade da prova foi testada recorrendo à análise fatorial confirmatória. Os dados obtidos com os vários indicadores de ajustamento considerados não são completamente congruentes. Os valores de CFI e TLI são muito elevados, indicando um ajustamento excelente do modelo aos dados. Todavia, o valor de qui quadrado e do RMSEA apontam em sentido oposto. Lembra-se que a PRP é uma prova de papel e lápis, na qual os alunos dispõem de um tempo limite para responder aos itens. Nos últimos itens verificam-se, por isso, sempre resultados omissos. Esta razão justifica que, para testar a unidimensionalidade da prova, se tenha optado por efetuar uma aplicação sem limite de tempo, cujos resultados se descreveram no Estudo 1. Neste estudo a amostra utilizada incluiu alunos dos vários anos de escolaridade a que a prova se destina. Em pesquisas posteriores será de considerar a realização de análises de invariância fatorial por ano, requerendo nos mesmos uma administração da prova sem tempo limite. É provável que deste modo se esclareçam melhor os resultados encontrados no Estudo 1. A maior parte das recomendações, muito genéricas, relativas ao número de casos por parâmetro não são válidas quando os itens são respondidos de forma binária, como é o caso da PRP, ou noutro tipo de categorias do tipo Likert (Finney & DiStefano, 2006). Ainda que esteja para além do propósito do presente estudo, pensamos que investigações futuras com esta escala poderiam pesquisar aspetos da estabilidade dos parâmetros locais da escala PRP a partir da metodologia *bootstrap*, bem como diferentes condições associadas ao número de itens e estrutura da solução (unidimensional versus multifatorial) tal como sugerido por L. K. Muthén e Muthén (2002).

Na PRP não foram encontradas diferenças entre sexos, corroborando os resultados encontrados por Lúcio (2008). Por sua vez Capovilla et al. (2006) obtiveram diferenças em alguns sub-testes. Os estudos efetuados numa ótica mais experimental e que têm contribuído para descrever os efeitos das características psicolinguísticas das palavras

não tomam o sexo como uma variável independente (ver a este a propósito a revisão recente efetuada por Lúcio & Pinheiro, 2011) o que pode indicar que não haverá razões para se supor a ocorrência de diferença nem uma justificação teórica para as mesmas. Investigações futuras poderão vir a esclarecer estes resultados.

O estudo da validade concorrente da prova foi delimitado recorrendo a vários procedimentos: a comparação entre anos de escolaridade; o cálculo do coeficiente de correlação entre os resultados na PRP e noutra prova de leitura; a avaliação do desempenho em leitura efetuada pelos professores. Os resultados encontrados apoiam a validade da prova.

Como era esperado, verificou-se que as médias nos resultados da PRP aumentam em função do ano de escolaridade e que as diferenças são significativas do ponto de vista estatístico. Os resultados observados indicaram que as diferenças de desempenho eram maiores do 1.º para o 2.º ano do que do 3.º para o 4.º ano. Apesar de o 1.º ano ser o ano da aprendizagem formal da leitura e da escrita, há especificidades, como os grafemas complexos, as vogais nasais e as regras contextuais cuja abordagem sistemática só ocorre durante o 2.º ano de escolaridade, pelo que o seu domínio só é consolidado ao longo deste ano. À medida que a escolaridade aumenta, a velocidade e a precisão no reconhecimento de palavras tende a uniformizar, na medida em que se espera que o processo de descodificação esteja automatizado no final do 4.º ano de escolaridade, com a prática a revelar-se como o principal fator de evolução posterior, como demonstraram os resultados de estudos comparativos realizados com estudantes do 4.º ao 8.º ano de escolaridade (Jenkins & Jewell, 1993; Kuhn & Stahl, 2003; Yovanoff et al., 2005).

No presente estudo considerou-se a precisão na leitura de palavras efetuada dentro de limites de tempo pré-definidos. Esta opção não permite, contudo, avaliar o tempo gasto na leitura de cada uma das palavras e, conseqüentemente, verificar as diferenças ocorridas em função da extensão de cada palavra e da classe de vizinhas candidatas à ativação. Embora o registro do tempo de reação e de resposta quer às palavras-alvo, quer às restantes palavras que compunham cada item, permitisse análises mais finas, a recolha destes elementos não é compatível com o uso desta prova como instrumento de *screening* em contextos escolares, dado exigir *software* específico. No entanto, consideramos que este aspecto não invalida as potencialidades desta prova, dado que as palavras-alvo dos itens da PRP são apenas palavras regulares e dado que não se verificaram diferenças de desempenho em função do número de sílabas das palavras ou em função do número de vizinhas ortográficas.

Apesar de o português ser uma língua bastante regular em termos de leitura, as dificuldades exibidas por muitos alunos desde o início da escolaridade não são suficientemente valorizadas, o que adia a referência e o diagnóstico de muitas crianças. Sendo a automatização

do reconhecimento de palavras decisiva para libertar recursos para a compreensão da leitura (Morais, 1997), a avaliação precoce de dificuldades a este nível mostra-se de especial importância. Recorde-se que a PRP é uma prova de *screening*, incluindo apenas palavras regulares. Atendendo às características da língua portuguesa em termos de leitura, face aos dados obtidos considera-se que a PRP é um instrumento fiável para a identificação precoce dos alunos que necessitem de intervenção ao nível da fluência e da precisão da leitura. Se as dificuldades são visíveis neste tipo de palavras, elas serão muito mais evidentes na leitura de palavras irregulares ou de pseudo palavras.

Referências

- Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory & Cognition*, *15*, 802-814.
- Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory & Cognition*, *18*, 234-254.
- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, *4*, 439-461.
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., ... Halgren, E. (2006). Top-down facilitation of visual word recognition. *Proceedings of the National Academy of Science, USA. Neuroscience*, *103*(2), 449-454.
- Becker, D. V., Goldinger, S. D., & Stone, G. O. (2006). Perception and recognition memory of words and words: Two-way mirror effects. *Memory & Cognition*, *34*, 1495-1511.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588-606.
- Blunch, N. (2008). *Introduction to structural equation modelling using SPSS and AMOS*. London: Sage.
- Brown, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equations models* (pp. 136-162). Newbury Park, CA: Sage.
- Capovilla, F., Varanda, C., & Capovilla, A. (2006). Teste de competência de leitura de palavras e pseudopalavras: Normalização e validação. *PSIC – Revista de Psicologia da Vetor Editora*, *7*(2), 47-59.
- Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of orthographic neighborhood in visual word recognition: Cross-tasks comparisons. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *23*, 857-871.
- Cohen, J. (1988). *Statistical power analyses for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). Hillsdale, NJ: Erlbaum.
- Cook, K. F., Kallen, M., & Amtmann, D. (2009). Having a fit: Impact of number of items and non-normality on tests of IRT's unidimensionality assumption. *Quality of Life Research*, *18*(4), 447-460.
- Feldt, L. S., & Kim, S. (2006). Testing the difference between two alpha coefficients with small samples of subjects and raters. *Educational and Psychological Measurement*, *66*(4), 589-600.
- Finney, S. J., & DiStefano, C. (2006). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. Mueller (Eds.), *Structural equation modeling. A second course* (pp. 269-314). Greenwich, CT: Information Age.
- Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology*, *33*(4), 465-495.
- Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*(3), 239-256.
- Gilmore, A., Croft, C., & Reid, N. (1981). *The Burt word reading test*. Wellington, New Zealand: New Zealand Council for Educational Research.
- Grainger, J. (2002). Structure et fonctionnement du lexique bilingue. In A. Florin & J. Morais (Eds.), *La Maîtrise du langage. Textes issus du XXVII^e Symposium de l'Association de Psychologie Scientifique de Langue Française* (pp. 41-59). Rennes, France: Presses Universitaires de Rennes.
- Grainger, J., & Segui, J. (1990). Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Perception & Psychophysics*, *47*, 191-198.
- Grainger, J., O'Reagan, K., Jacobs, A., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effects. *Perception & Psychophysics*, *45*, 189-195.
- Hoge, R. D., & Caladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, *59*, 297-313.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.
- Jenkins, J. R., & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and maze. *Exceptional Children*, *59*(5), 421-432.
- Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology*, *95*(4), 719-729.
- Johnson, D. D. (1973). Sex differences in reading across cultures. *Reading Research Quarterly*, *9*(1), 67-86.
- Justi, F. R. R., & Pinheiro, A. M. V. (2006). O efeito de vizinhança ortográfica no Português do Brasil: Acesso lexical ou processamento estratégico. *Revista Interamericana de Psicologia*, *40*(3), 275-288.
- Kline, R. (2005). *Principles and practice of structural equation modeling*. New York: The Guilford Press.
- Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology*, *95*(1), 3-21.
- Lúcio, P. S. (2008). *Investigação psicométrica de uma tarefa de leitura em voz alta de palavras isoladas* (Dissertação de mestrado, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil).
- Lúcio, P. S., & Pinheiro, A. M. V. (2011). Vinte anos de estudo sobre o reconhecimento de palavras em crianças falantes do português: Uma revisão da literatura. *Psicologia: Reflexão e Crítica*, *24*(1), 170-179.
- Mather, N., Hammill, D. D., Allen, E. A. L., & Roberts, R. (2004). *Test of silent word reading fluency: Examiner's manual*. Austin, TX: PRO-ED.

- McGuiness, D. (2005). *Language development and learning to read*. Cambridge, MA: MIT Press.
- Milanowski, A. T., Kimball, S. M., & White, B. (2004). *The relationship between standards-based teacher evaluation scores and student achievement: Replication and extensions at three sites*. Madison, WI: University of Wisconsin. Retrieved in October 10, 2010, from http://cpre.wceruw.org/papers/3site_long_TE_SA_AERA04TE.pdf
- Morais, J. (1997). *A arte de ler: Psicologia cognitiva da leitura*. Lisboa, Portugal: Edições Cosmos.
- Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, 76, 165-178.
- Morton, J. (1982). Desintegrating the lexicon. In J. Mehler, E. C. T. Walker, & M. Garret (Eds.), *Perspectives in mental representations* (pp. 89-109). Hillsdale, NJ: Lawrence Erlbaum.
- Motl, R., & DiStefano, C. (2002). Longitudinal invariance of self-esteem and method effects associated with negatively worded items. *Structural Equation Modeling*, 9, 562-578.
- Muthén, B. O., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes* (Technical report). Los Angeles, CA: University of California.
- Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Author.
- Muthén, L. K. & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 4, 599-620.
- Nikaedo, C. C., Kuriyama, C., & Macedo, E. C. (2007). Avaliação longitudinal de leitura e escrita com testes de diferentes pressupostos teóricos. *Psic*, 8, 65-73.
- Nunnally, J. (1978). *Psychometric theory*. New York: MacGraw-Hill.
- Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation-verification model of letter and word recognition: The word superiority effect. *Psychological Review*, 89, 573-594.
- Perea, J. (2008). The effect of neighborhood frequency in reading: Evidence with transposed-letter neighbors. *Cognition*, 108(1), 290-300.
- Perea, J., & Rosa, E. (2000). The effects of orthographic neighbourhood in reading and laboratory word identification tasks: A review. *Psicológica*, 21, 327-340.
- Perea, M., Carreiras, M., & Grainger, J. (2004). Blocking by word frequency and neighborhood density in visual word recognition: A task-specific response criteria account. *Memory & Cognition*, 32, 1090-1102.
- Rumelhart, E. D., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception, Part 2: The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, 89, 69-94.
- Segui, J. (1991). La reconnaissance visuelle de mots. In R. Kolinsky, J. Morais, & J. Segui (Eds.), *La reconnaissance des mots dans les diferentes modalités sensorielles* (pp. 99-117). Paris: Presses Universitaires de France.
- Sim-Sim, I., & Viana, F. L. (2007). *Para a avaliação do desempenho de leitura*. Lisboa, Portugal: Gabinete de Estatística e Planeamento da Educação, Ministério da Educação.
- Snowling, M., & Hulme, C. (2005). *The science of reading: A handbook*. London: Blackwell.
- Stein, L. M. (2008). *Teste de Desempenho Escolar*. São Paulo, SP: Casa do Psicólogo.
- Sucena, A., & Castro, S. L. (2008). *Aprender a ler e avaliar a leitura. O TIL – Teste de Idade de Leitura*. Coimbra, Portugal: Livraria Almedina.
- Viana, F. L., & Ribeiro, I. (2006). *Avaliar Leitura. Apresentação de uma Prova de Leitura de Palavras*. Trabalho apresentado no VI Encontro Nacional (IV Internacional) de Investigação em Leitura, Literatura Infantil e Ilustração, Braga, Portugal.
- Viana, F. L., & Ribeiro, I. (2007, 25-28 jul.). *P.R.P. - Prova de Reconhecimento de Palavras. Contributos para a validação de um instrumento de avaliação de leitura*. Trabalho apresentado no III Congresso Brasileiro de Avaliação Psicológica: Formas e Contextos, João Pessoa, PB, Brasil.
- Walla, P., Hufnagl, B., Lindinger, G., Deecke, L., & Lang, W. (2001). Physiological evidence of gender differences in word recognition: A magnetoencephalographic (MEG) study. *Brain Research. Cognitive Brain Research*, 12(1), 49-54.
- White, B. (2004). *The relationship between teacher evaluation scores and student achievement: Evidence from Coventry, RI*. Madison, WI: University of Wisconsin. Retrieved October 1, 2010, from <http://www.smhc-cpre.org/download/74/>
- Yovanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice*, 24(3), 4-12.
- Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* (Unpublished doctoral dissertation). University of California, Los Angeles, CA, USA.

Recebido: 26/12/2011
1ª revisão: 29/03/2012
Aceite final: 30/03/2012