

Desenvolvimento de um teste computadorizado de avaliação da ortografia

Débora Cecilio Fernandes¹

Universidade Estadual do Centro-Oeste, Guarapuava-PR, Brasil

Fermino Fernandes Sisto

Universidade Estadual de Campinas, Campinas-SP, Brasil

RESUMO

Este estudo teve como objetivo desenvolver uma nova versão computadorizada da Escala de Avaliação de Reconhecimento de Palavras. A escala foi aplicada inicialmente, em 2001, em 338 alunos de segundo a quinto ano de 3 escolas do interior de São Paulo. A média de idade das crianças foi de 8,81 anos ($DP=0,42$). Em 2013, a versão computadorizada foi aplicada em 269 alunos de segundo a quinto ano, com média de idade de 8,90 ($DP=1,06$). Foram eliminados 116 itens, com base no índice de dificuldade, níveis de dificuldades redundantes e indicadores de ajuste e dificuldade linguística, originando uma escala de 55 itens. A análise de Rasch evidenciou melhor ajuste e a ANOVA indicou diferenças significativas entre as médias das crianças de segundo a quinto ano. Os resultados foram interpretados como evidências de validade da nova versão do instrumento; no entanto, outros estudos psicométricos ainda são necessários.

Palavras-chave: escrita; ortografia; psicométrica.

ABSTRACT – Development of a computerized orthography assessment test

This study aimed to develop a new computerized version of the Word Recognition Assessment Scale. The scale was initially administered in 2001 to 338 students from second to fifth grades of three schools from São Paulo's countryside. Children's mean age was 8.81 years old ($SD=.42$). In 2013, the computerized version was administered to 269 students from second to fifth grades, whose mean age was 8.90 ($SD=1.06$). Some 116 items were eliminated, based on difficulty index, redundant difficulty levels, adjustment indexes and linguistic difficulty, originating a scale composed of 55 items. Rasch analysis evidenced a better adjustment, and ANOVA indicated significant differences among children from second to fifth grades. Results were interpreted as validity evidences for the new version of the instrument. Nevertheless, further psychometric studies are necessary.

Keywords: writing; orthography; psychometrics.

RESUMEN – Desarrollo de un test informatizado de evaluación de la ortografía

El presente estudio tuvo como objetivo desarrollar una nueva versión informatizada de la Escala de Evaluación del Reconocimiento de Palabras. Inicialmente, en 2001, se aplicó la escala a 338 alumnos de segundo a quinto curso de tres escuelas del interior de São Paulo. La edad media fue 8,81 años ($DE=0,42$). En 2013, la versión informatizada se aplicó en 269 alumnos de segundo a quinto curso, con edad media de 8,90 ($DE=1,06$). Se eliminaron 116 ítems, con base en los índices de dificultad, niveles de dificultad redundantes, indicadores de ajuste y dificultades lingüísticas, produciendo una escala de 55 ítems. El análisis de Rasch evidenció un mejor ajuste y el ANOVA indicó diferencias significativas entre las medias de los niños del segundo al quinto curso. Los resultados se interpretaron como evidencias de validez de la nueva versión del instrumento. Indicaron que la nueva versión fue más discriminativa que el instrumento original y que se ajustó mejor, sin embargo, más estudios psicométricos todavía se hacen necesarios.

Palabras clave: escritura; ortografía; psicometría.

Ao ingressar na escola, a criança descobre um novo mundo repleto de diferentes conhecimentos e regras que deve aprender. Nesse contexto se enquadra a aquisição das habilidades escolares como a leitura e a escrita, que pode ser abordada a partir de distintos pontos de vista, como a educação formal, o desenvolvimento das capacidades cognitivas ou da aquisição de habilidades e processos cognitivos durante a escolarização, entre outros.

A escrita pode ser compreendida como a expressão da linguagem oral dos sinais criados pelo ser humano (Cagliari, 2010). Estudos sobre a escrita são normalmente divididos em duas grandes áreas. A primeira delas se centra nos processos de codificação e foca os diferentes sistemas de escrita, sua natureza, os aspectos notacionais e o desenvolvimento da aquisição do sistema alfabético, que aborda as relações entre a linguagem oral e a escrita.

¹Endereço para correspondência: Rua Frei Caneca, 3159, Santa Cruz, 85015-220, Guarapuava-PR, Brasil. E-mail: debora.cecilio@gmail.com

A segunda grande área de estudos da escrita se relaciona com a sua composição enquanto texto, com unidades linguísticas maiores do que palavras e frases. Seu foco é no discurso, no texto e em seus gêneros (Báez, 2013).

Por sua vez, a investigação da leitura pode ser considerada a partir dos denominados micro e macroprocessos (Urquijo, 2010). De modo geral, os microprocessos estão relacionados à decodificação do texto; por exemplo, o reconhecimento de letras, a construção silábica, a decodificação de palavras, entre outros. A decodificação na leitura refere-se à conversão das letras do texto escrito em seus sons correspondentes, o que possibilita a formação do código grafêmico e do código fonêmico (Urquijo, 2010). Os macroprocessos se referem às operações de nível superior, como a compreensão da leitura. A compreensão da leitura é uma tarefa complexa que requer o funcionamento de processos e habilidades cognitivas, a integração da informação de sentenças e ideias em um texto, o monitoramento e o conhecimento sobre a estrutura do texto (Cain, Bryant, & Oakhill, 2004), assim como a formação de um modelo mental sobre o texto (Kintsch, & Dijk, 1978).

Considera-se que a leitura e a escrita são processos integrados e relacionados, ou seja, a aquisição de um deles depende da aquisição do outro. O modelo teórico elaborado por Berninger et al. (por exemplo, Berninger, Abbott, Abbott, Graham, & Richard, 2002; Jones, Abbott, & Berninger, 2014) considera a aquisição da leitura e escrita como processos interdependentes, em um contexto de desenvolvimento cognitivo. Berninger et al. (2002) desenvolveram um sistema cognitivo integrado que vincula os processos de escrita e os de leitura entre si e com os diferentes processos cognitivos necessários para o desenvolvimento das habilidades escolares. Para o presente estudo, o foco foi o modelo específico de escrita denominado *Not-So-Simple View of Writing* (Berninger et al., 2002; Jones et al., 2014), que inclui quatro componentes: funções executivas, a memória de trabalho, a geração de texto e a transcrição.

Brevemente, as funções executivas são um conjunto de habilidades cognitivas responsáveis pelo engajamento em comportamentos orientados aos objetivos definidos, pela realização de ações voluntárias, independentes e auto-organizadas (Diamond, 2013). Dentre os diferentes processos executivos está a memória de trabalho, que se refere a um sistema ativo de armazenamento de informações de curto prazo e de capacidade limitada que possibilita a manipulação dessas informações (Baddeley, 2012). Esse sistema mnemônico recupera informações necessárias dos sistemas de memória de longo prazo para seu uso no momento. Empregando o modelo de memória de trabalho de Baddeley (2012), Brooks, Berninger e Abbott (2011) e Berninger, Richards e Abbott (2015) consideram que o papel da memória de trabalho é integrar os processos de geração da escrita, as funções executivas de forma geral e a transcrição, que se refere à recuperação

de símbolos ortográficos da memória de longo prazo em combinação com as respostas motoras requeridas para produzir esses símbolos.

A geração da escrita, por sua vez, é compreendida como a transformação de ideias atualizadas ou mantidas na memória de trabalho em linguagem (Berninger et al., 2002). Como continuidade dessa transformação, entra em cena o processo de transcrição, que integra a grafia (movimentos motores), mediante a escrita à mão ou a digitação, e a ortografia. Dessa forma, a transcrição é o processo que transforma as representações mentais em um texto escrito (Báez, 2013; Berninger et al., 2002). O enfoque do presente estudo, sem desconsiderar a leitura e a complexidade dos processos cognitivos envolvidos no modelo *Not-So-Simple View of Writing*, é a escrita. Especificamente, o interesse maior recai sobre o processo de transcrição.

A capacidade de transcrição das crianças se desenvolve conforme elas entram em contato com a ortografia e aprendem como usá-la, praticam e melhoram suas habilidades motoras finas e refinam a coordenação entre a lembrança da ortografia e da forma escrita das palavras e os movimentos motores necessários para a escrita das letras e palavras. A aprendizagem de cada um desses processos é demorada e laboriosa, dificultada ainda mais pela necessidade de integração de ambas as habilidades (Berninger et al., 2002).

Nesse sentido, ao levar em consideração a melhoria gradativa esperada do processo de transcrição, os erros fazem parte da aprendizagem. No entanto, em determinados momentos da alfabetização, espera-se que alguns erros deixem de ocorrer. Assim, a avaliação da escrita e da ortografia das crianças é uma forma de identificar se estão adquirindo o conhecimento da escrita ortográfica esperado. Esse conhecimento, inicialmente básico, é responsável pelo desenvolvimento de habilidades de escrita mais complexas, como a escrita fluente e correta de orações, parágrafos e, futuramente, textos completos (Báez, 2013; Berninger et al., 2002).

Para tal, é fundamental que os instrumentos de avaliação dos diferentes processos ou subprocessos relacionados à escrita estejam disponíveis para psicólogos, psicopedagogos e professores. Essas avaliações devem passar por processos de validação, devem ser precisas e, preferencialmente, de fácil aplicação. A Escala de Avaliação de Reconhecimento de Palavras, inicialmente construída por Kingeski e Sisto (2004), contém 221 itens e tem como objetivo avaliar se as crianças reconhecem a escrita ortográfica de palavras ouvidas. Nesse instrumento, após escutar uma palavra, são apresentadas visualmente três opções de escrita correspondentes à palavra ouvida. Apenas uma dessas opções representa a escrita correta da palavra. Assim, trata-se de uma avaliação do reconhecimento ortográfico.

A vantagem desse formato do instrumento se refere ao isolamento de uma das habilidades necessárias para a transcrição, ou seja, o conhecimento ortográfico é

avaliado mediante o seu reconhecimento, e não é necessário que a criança mostre esse conhecimento por meio da escrita, como ocorre com o ditado. Avaliar apenas o conhecimento que a criança possui da escrita correta da língua é interessante nos anos iniciais da alfabetização, pois essa avaliação não se confunde com a habilidade motora que a criança tenha ou não para a escrita. Como inicialmente os movimentos da escrita são novos e laboriosos para os alunos, eles devem dedicar muitos recursos motores e cognitivos para essa atividade. Nesse sentido, a maior parte da atenção se volta para os próprios movimentos da escrita das letras e palavras e para a recuperação da forma visual de cada letra ou palavra. Por isso, um instrumento capaz de separar a habilidade motora do conhecimento ortográfico é importante para uma avaliação mais precisa da ortografia dos escolares.

O estudo dessa primeira versão de 221 itens mostrou que a prova não diferenciou significativamente entre o desempenho de crianças de 4^o e 5^o ano (Kingski, & Sisto, 2004). Esse resultado foi preocupante, pois se espera que crianças do quinto ano tenham melhor conhecimento da ortografia da língua portuguesa que as do quarto ano, e que a prova de ortografia seja capaz de diferenciar esses alunos. Assim, surgiu a necessidade de uma análise pormenorizada dos itens do instrumento e um estudo de validação cruzada. A análise dos itens deve auxiliar a acrescentar, alterar ou excluir itens de um teste em função de análises quantitativas e qualitativas, e a validação cruzada é usada com o intuito de averiguar se as estatísticas dos itens se mantêm quando aplicadas em diferentes amostras, até que se obtenha um conjunto satisfatório de itens (Urbina, 2007). Dessa forma, o objetivo deste estudo foi aprimorar e desenvolver uma nova versão da Escala de Avaliação de Reconhecimento de Palavras, por meio de um estudo de validação cruzada.

Método

Este estudo inclui duas etapas de coleta de dados. A primeira foi realizada em 2001, cujos dados foram retomados e foram realizadas novas análises. A partir delas, elaborou-se uma segunda versão da Escala que, após nova coleta em 2013, foi submetida à análise pelo modelo de Rasch.

Participantes

Em 2001, 338 crianças e adolescentes do Estado de São Paulo participaram da coleta. Os participantes estavam regularmente matriculados nos segundos, terceiros, quartos e quintos anos do ensino fundamental de duas escolas públicas e de uma escola particular. Todas as crianças apresentaram o Termo de Consentimento Livre e Esclarecido (TCLE) assinado por um dos pais e responsáveis. Em relação aos dados dos participantes, dez pessoas não informaram a idade e uma pessoa do terceiro ano escolar não informou o sexo. A Tabela 1 mostra a caracterização dos participantes, em razão do ano escolar e das idades.

Observa-se que o quinto ano possui número mais elevado de participantes, seguido do quarto ano. Nota-se que, de forma geral, há uma adequação da idade dos participantes aos anos escolares. A média geral da idade dos participantes foi de 8,81 anos, com desvio padrão de 1,33. Em relação ao sexo das pessoas, 145 (42,9%) eram do sexo masculino.

Na Tabela 1, encontram-se ainda as informações referentes aos participantes da coleta de 2013. Dessa segunda coleta participaram 269 alunos de segundo a quinto ano do ensino fundamental I. Nota-se que a maior quantidade de alunos estava cursando o terceiro ano escolar, seguido do segundo ano. Novamente, é possível observar que, de forma geral, há uma adequação da idade em relação ao ano escolar. A média geral de idade foi de 8,90 ($DP=1,06$) anos e 140 alunos eram do sexo masculino (52%). Ressalta-se que todos os participantes apresentaram o TCLE assinado por um dos pais ou responsáveis.

Instrumento

Escala de Avaliação do Reconhecimento de Palavras – versão inicial

O instrumento utilizado neste momento da pesquisa foi a Escala de Avaliação do Reconhecimento de Palavras, cuja versão inicial foi desenvolvida por Kingski e Sisto (2004). Para a construção dessa primeira versão, foram selecionados apostilas e livros

Tabela 1
Distribuição dos participantes da coleta de 2001 e 2013 em razão da idade e do ano escolar

Idade	Ano escolar				Total	%
	2	3	4	5		
2001						
6	5	-	-	-	5	1,5
7	61	2	-	-	63	19,2
8	8	57	3	-	68	20,7
9	-	20	61	-	81	24,7
10	-	1	17	58	76	23,2
11	-	-	1	33	34	10,4
12	-	-	1	-	1	0,3
Total	74	80	83	91	328	100
2013						
7	19	0	0	0	19	7,1
8	58	33	0	0	91	33,8
9	1	50	21	0	72	26,8
10	0	3	36	32	71	26,4
11	0	0	6	10	16	5,9
Total	78	86	63	42	269	100

didáticos de grande representatividade no Brasil. Foram construídos 221 itens com base em palavras escolhidas aleatoriamente do material selecionado. As palavras foram classificadas de acordo com a presença ou ausência de dificuldades da língua portuguesa, conforme Sisto (2001). Assim, foram consideradas palavras com dificuldades linguísticas que continham encontros consonantais, dígrafos, sílabas compostas, sílabas complexas, ou as que eram polissílabas. Essa análise das dificuldades das palavras mostrou que 165 delas eram ligadas a dificuldades de escrita (74,66%).

Cada item era composto por três opções de respostas, uma das quais era a palavra escrita corretamente. Os erros ortográficos variaram conforme os erros mais comuns da língua portuguesa (Sisto, 2001) e incluíam trocas vocálicas (por exemplo, em palavra terminada em *o*, em uma das opções de resposta equivocada a palavra termina em *u*), trocas de consoantes com a mesma sonoridade (por exemplo, *ç*, *c* e *ss*), troca de *n* por *m*, entre outras, além de erros de acentuação. A escolha da resposta correta recebia um ponto e a escolha das opções erradas recebia zero, o que formou uma escala de 0 a 221 pontos.

Escala de Avaliação do Reconhecimento de Palavras – versão computadorizada

A Escala de Avaliação do Reconhecimento de Palavras – versão computadorizada (Fernandes, Sisto, & Murarolli, 2014) foi desenvolvida em função das análises realizadas com os dados da coleta de 2001 da versão original do instrumento com 221 itens. Como resultado, obteve-se uma escala com 55 itens, que foi implementada em uma plataforma computadorizada. Para tal, foi utilizado o *software LiveCode* (versão 6.1). A versão computadorizada possui instruções pronunciadas por um profissional de rádio com ótima dicção e sem sotaque, que foram gravadas no formato *.mp3* e convertidas para o formato *.aif* com o uso do *software AIFF MP3 Convert*.

A título de exemplo de apresentação de item, é exibido um botão na tela do computador para que a criança aperte e escute uma palavra. Após a emissão do som da palavra, a tela muda e aparecem, ao mesmo tempo, três palavras. Essas três palavras representam um item. Podem aparecer as palavras *alem*, *alén* e *além*, que compõem um item, ou ainda, *poduz*, *produz* e *produs*, *ambienti*, *ambiente*, *geraumente*, *geralmente* e *geralmente*, etc. A criança deve clicar na palavra que considerar escrita corretamente. Por cada acerto, recebe-se um ponto.

Procedimento

A coleta de dados de 2001 foi inicialmente aprovada pelo Comitê de Ética em Pesquisa da Universidade São Francisco, que considerou que a pesquisa respeitava os preceitos éticos da Resolução 196/96, do Conselho Nacional de Saúde (CNS). Após a sua aprovação, foi realizada a aplicação do instrumento em três escolas. As aplicações foram coletivas e demoraram aproximadamente

uma hora. O pesquisador explicou as instruções da prova para as crianças e deu início à aplicação coletiva em sala de aula. A instrução do pesquisador foi “Depois de ouvir a palavra, faça um X na palavra escrita corretamente”. A aplicação consistiu na leitura de cada palavra. Após escutar as palavras, as crianças deveriam escolher a palavra correspondente escrita corretamente.

A coleta de 2013 foi aprovada pelo Comitê de Ética em Pesquisa da Universidade do Vale do Sapucaí (protocolo número: CAEE 22821613.0.0000.5102). Nesse momento, o instrumento havia sido reduzido para 55 itens e estava disponível em versão computadorizada. A aplicação da Escala foi acompanhada por uma das pesquisadoras, na sala de informática da escola. A pesquisadora instalou o programa nos computadores, o que permitiu uma coleta coletiva. As crianças ouviram com fones de ouvido as instruções ditadas pelo próprio *software*. As instruções explicavam que as crianças deveriam apertar o botão indicado para ouvir a palavra, e quando aparecessem as três palavras escritas na tela, deveriam clicar na que considerassem estar escrita corretamente. Tais instruções eram seguidas de exemplos e, posteriormente, iniciava-se a prova. As crianças responderam, sozinhas e sem dificuldades, aos itens, processo que durou cerca de 20 minutos. Ressalta-se que, em ambas as coletas, todos os procedimentos éticos foram seguidos conforme o constante na Declaração de Helsinki (1975, revisada em 1983) e as indicações das Resoluções nº 196, de 10 de outubro de 1996, e nº 251, de 07 de agosto de 1997, do CNS.

Análise de dados e critérios adotados para a redução do número de itens

Para cumprir com o objetivo da pesquisa, foi calculado o índice de dificuldade de todos os itens da escala para segundo, terceiro, quarto e quinto anos escolares. Esse índice é a proporção de acerto das palavras para o número de participantes de cada ano, o que permite averiguar em quais anos escolares as palavras obtiveram mais acertos. Espera-se que o número de acertos aumente conforme o ano escolar, o que é representado pelo aumento do índice de dificuldade. Nesse primeiro momento, foram usados como critérios para a redução de itens a inversão do índice de dificuldade com relação aos anos escolares (por exemplo, quando uma palavra é mais difícil para o terceiro ano que para o segundo), a igualdade entre os índices de dificuldade de dois anos consecutivos e a pouca discriminação entre, principalmente, quartos e quintos anos.

Após a eliminação dos itens com base nesses critérios, foi realizada a análise pelo modelo de Rasch (1960) para estimar o nível de dificuldade dos itens e localizar os itens e as pessoas na escala *logit*. Optou-se pelo uso desse modelo, apesar de não haver estimação dos valores de discriminação dos itens, devido à amostra relativamente pequena de cada etapa da presente pesquisa. Enquanto o modelo de Rasch estima os parâmetros com amostras de

cem participantes, outros modelos da Teoria de Resposta ao Item requerem uma amostra muito mais ampla.

Foram calculados também os indicadores de ajuste (*infit* e *outfit*) dos itens e dos participantes ao modelo. Os valores de *infit* e *outfit* são a média das diferenças entre as respostas observadas e as esperadas, por isso indicam se as pessoas responderam aos itens de acordo com o esperado pelo modelo e se os itens foram respondidos conforme previsão do modelo de Rasch. Nesse sentido, Linacre (2002) considera que valores mais próximos de 1,0 são ideais. Os valores entre 0,5 e 1,5 são interpretados como adequados e por isso os itens ou pessoas com esses valores são mantidos. Valores de 1,5 a 2,0 são considerados como divergências moderadas, mas não degradam as medidas. Quando os indicadores oferecem valores superiores a 2,0, os itens ou pessoas apresentam desajuste aberrante e devem ser revistos, excluídos ou modificados. Finalmente, valores inferiores a 0,5 são menos produtivos, mas podem ser mantidos.

Os itens com dificuldades iguais ou próximas foram considerados redundantes para a avaliação da habilidade em questão (Embretson, & Reise, 2000) e por isso puderam ser eliminados da escala sem prejudicar a precisão da medida. Para a decisão de quais itens com dificuldades parecidas deveriam ser mantidos ou eliminados, foram mantidos aqueles com melhores ajustes (Linacre, 2002) e cujas palavras possuísem dificuldades linguísticas não contempladas em outros itens.

Posteriormente, os itens restantes foram submetidos a uma nova análise pelo modelo de Rasch, para averiguar a sua adequação a uma escala por meio dos indicadores de ajuste e obter os valores de fidedignidade e dificuldade dos itens. Para as análises descritas foram utilizados os *softwares* SPSS (versão IBM-21) e Winsteps (versão 37004; Linacre, 2009). A seguir, são apresentados os resultados da pesquisa.

Resultados

Inicialmente, foi calculado o índice de dificuldade de cada palavra do instrumento em função do ano escolar. A Tabela 2 mostra exemplos de palavras que foram mantidas e/ou eliminadas com base nos critérios apresentados. Observa-se, também, que os índices de dificuldade das palavras mantidas aumentaram gradativamente com o incremento dos anos escolares. As palavras eliminadas estão em negrito e são apresentados os índices de dificuldade que justificaram a eliminação desses itens. Por exemplo, as palavras aberta, obras e Lúcia apresentaram uma inversão da dificuldade em função do ano escolar. As palavras aberta e obras foram mais fáceis para o segundo que para o terceiro ano. A palavra Lúcia foi mais fácil para o terceiro que para o quarto ano. A palavra carro foi igualmente fácil para o quarto e quinto anos. As palavras mas, beija-flor, continuou e pés obtiveram índices muito próximos do quarto e quinto anos escolares e por

isso foram eliminadas. Do total de 221 palavras, 95 foram excluídas com base nesses critérios, restando então 126.

Após a eliminação, o novo conjunto de 126 itens foi submetido à análise pelo modelo Rasch, com vistas a prosseguir com a possível exclusão de mais itens e estudar sua adequação a uma escala. Uma síntese dos resultados de ajuste pode ser visualizada na Tabela 3.

Os dados concernentes aos 126 itens indicaram que, no geral, a maioria deles foi respondida de acordo com o padrão esperado. Em uma análise mais detalhada do *infit*, observou-se que poucos apresentaram desajuste moderado ou severo (Linacre, 2002). Dois itens mostraram desajuste moderado com *infit* entre 1,50 e 2,0.

Em relação ao *outfit*, seu valor médio sugeriu uma boa adequação ao modelo. Foram encontrados 6 itens com valores discrepantes, 4 deles superiores a 1,50, mas inferiores a 2,0 e 2 itens superiores a 2,0. Esses itens tiveram acertos não esperados em razão da habilidade das pessoas em termos de conhecimento ortográfico.

A Figura 1 mostra os mapas do construto das versões de 126 e 55 itens da prova de ortografia, que indicam a distribuição das pessoas e dos itens por *logits*. O lado esquerdo da Figura 1 mostra o mapa do construto dos 126 itens e o lado direito, o mapa dos 55 itens. Ao lado esquerdo de cada mapa estão distribuídas as crianças, em função de sua habilidade na escala *logit*, e ao lado direito estão os itens, localizados de acordo com sua dificuldade. A média da dificuldade dos itens é convencionalmente estabelecida pelo modelo como zero e representada no mapa pela letra *M*. A letra *M* ao lado esquerdo dos mapas indica a média de habilidade das crianças, enquanto a letra *M* do lado direito dos mapas representa a média de dificuldade dos itens.

Tabela 2
Exemplos de palavras mantidas e eliminadas da Escala de Reconhecimento de Palavras

Palavras	2º ano	3º ano	4º ano	5º ano
Mantidas				
Além	0,11	0,16	0,31	0,42
Médico	0,13	0,26	0,27	0,34
Madrugada	0,19	0,22	0,27	0,32
Personagens	0,18	0,23	0,26	0,33
Eliminadas				
Aberta	0,21	0,09	0,29	0,41
Obras	0,22	0,20	0,27	0,31
Lúcia	0,16	0,24	0,22	0,38
Carro	0,21	0,25	0,25	0,29
Mas	0,19	0,24	0,28	0,29
Beija-flor	0,19	0,23	0,28	0,31
Continuou	0,18	0,25	0,27	0,30
Pés	0,17	0,26	0,27	0,30

A escala do mapa dos 126 itens varia de +6 a -5. Cada # significa 3 crianças e cada . representa 1 ou 2 crianças.

Em relação ao mapa dos 126 itens, nota-se que a habilidade média das crianças está muito acima da dificuldade média dos itens. Ainda, os itens se agruparam entre os valores +3,63 e -2,02 da escala *logit*, mostrando uma variabilidade de 5,65, o que pode ser considerado uma boa extensão. Por sua vez, as pessoas se localizaram entre 5,31 e -6,39 em uma extensão de medida de 11,70 *logits*.

Para dar continuidade à eliminação de itens, levouse em consideração o seu nível de dificuldade. Foram levantados duplas ou conjuntos de itens com dificuldades iguais ou muito próximas; para a sua exclusão, foram considerados também os itens vizinhos com a ajuda visual do mapa apresentado na Figura 1. Por exemplo, o mapa mostra que há uma grande aglomeração de itens entre -1 e +1 na escala *logit*. Quando necessário, para auxiliar a escolha do item que seria eliminado, também foi utilizado como critério o ajuste dos itens. Nesses casos, foram eliminados os que mais se afastavam do valor ideal de ajuste de 1,00. Ainda, quando esses critérios não eram suficientes para tomar a decisão de eliminação, foi considerada a ortografia da palavra, ou seja, foram mantidas as palavras que representassem alguma complexidade ortográfica que não estivesse contemplada em outras palavras do teste. A Tabela 4 mostra alguns exemplos de duplas ou conjuntos de itens com estimativas de dificuldades muito parecidas, assim como seus ajustes.

Na Tabela 4, os itens em negrito foram eliminados do teste. Por exemplo, as palavras receita e teatro possuem a mesma dificuldade; assim, optou-se pela eliminação de teatro, pois seus indicadores de ajuste são muito menos adequados que os da palavra receita. Entre as palavras de mesma dificuldade amarela e acesso, decidiu-se por manter amarela, já que seus indicadores de ajuste são mais adequados que os de acesso. Por sua vez, entre as palavras jogara e material, houve uma preferência pela palavra material, novamente devido ao ajuste mais adequado dessa palavra. A escolha pela palavra descubra em detrimento da palavra extraveiculares se deu pelo mesmo motivo dos itens anteriores.

Finalmente, as palavras clareira, adivinhe, decorrer e exemplo possuem dificuldades muito próximas e fazem parte de uma aglomeração de itens. Optou-se pela eliminação das palavras adivinhe e decorrer em vez das palavras clareira ou exemplo, em função de suas dificuldades linguísticas. Por exemplo, a dificuldade da palavra eliminada adivinhe é o *nh*, que está presente em outras palavras da escala. No caso da palavra eliminada decorrer, a dificuldade seria o uso do *rr*, que também está presente em outras palavras. Por isso, decidiu-se por manter as palavras clareira e exemplo. Clareira contém *cl* e *ei* (normalmente, há uma omissão do som da letra *i* na pronúncia) como dificuldades, e a palavra exemplo contém o uso da letra *x* e de *pl*. Com base nesses critérios, foram eliminados mais 71 itens e, assim, chegou-se a um instrumento com 55 itens.

Após a seleção das palavras para compor o instrumento final, foi realizada uma nova coleta com dados mais atuais, para verificar se a versão final se ajustaria a uma escala *logit* pelo modelo de Rasch e se seria possível diferenciar os acertos de crianças dos distintos anos escolares, visto que as 221 palavras não apresentaram distinção entre o 4º e 5º anos. A Tabela 3 mostra os resultados dessa nova análise pelo modelo de Rasch da Escala de Avaliação do Reconhecimento de Palavras (versão computadorizada) composta por 55 itens. Observa-se que os itens se ajustaram bem ao modelo proposto. Poucos itens apresentaram desajuste leve e moderado em relação ao *oufit*.

Ainda em relação à análise de Rasch, na Figura 1, observa-se o mapa do construto da escala com 55 itens. A escala oscila entre +5 e -3 *logits*. Nesse caso, cada # representa duas crianças e cada . representa uma. Podem ser observadas as palavras da escala e sua localização no mapa. As palavras mais difíceis de serem reconhecidas ortograficamente foram posição, importante e proparoxítona. As mais fáceis foram estiagem, palavra, errado e cinzenta.

Após a nova coleta de dados, realizou-se uma análise de variância (ANOVA) com os valores dos parâmetros de habilidade das crianças obtidos pelo modelo de Rasch para averiguar se diferem em função do ano escolar. A análise indicou que $F(338,3)=40,266, p<0,001$, o que mostra que as diferenças não eram casuais. O resultado do teste *a posteriori*

Tabela 3
Parâmetros de ajuste dos itens das versões da escala de 126 e 55 itens

Parâmetros	Escala de 126 itens		Escala de 55 itens	
	Infit	Outfit	Infit	Outfit
Média	0,99	0,97	1,00	1,00
Desvio padrão	0,16	0,58	0,12	0,20
Máximo	1,82	6,59	1,25	2,48
Mínimo	0,71	0,41	0,77	0,61
Desajuste moderado (1,5-2,0)	2 (1,58%)	4 (3,17%)	0	6 (3,3%)
Desajuste alto (>2,0)	0 (0,00%)	2 (1,58%)	0 (0,0%)	0 (0,0%)

de Tukey indicou que a média das crianças do 2º ano foi de -0,45, significativamente diferente de todas as outras médias (com valor de $p < 0,001$ para todas as comparações). A média do 3º ano foi de 0,35 e se diferenciou significativamente das demais médias (p para a comparação entre o 3º e 4º anos foi 0,001, e para as outras comparações, foi $p < 0,001$). A média do 4º ano foi de 1,07, significativamente distinta das demais

(para a comparação entre 4º e 5º anos; $p = 0,48$). Finalmente, a média dos alunos de 5º ano foi de 1,64, significativamente superior às médias dos outros anos escolares. O tamanho do efeito foi de 0,488, considerado grande. Assim, a análise formou quatro subconjuntos diferentes estatisticamente, indicando que houve diferença significativa entre as médias de todos os anos escolares.

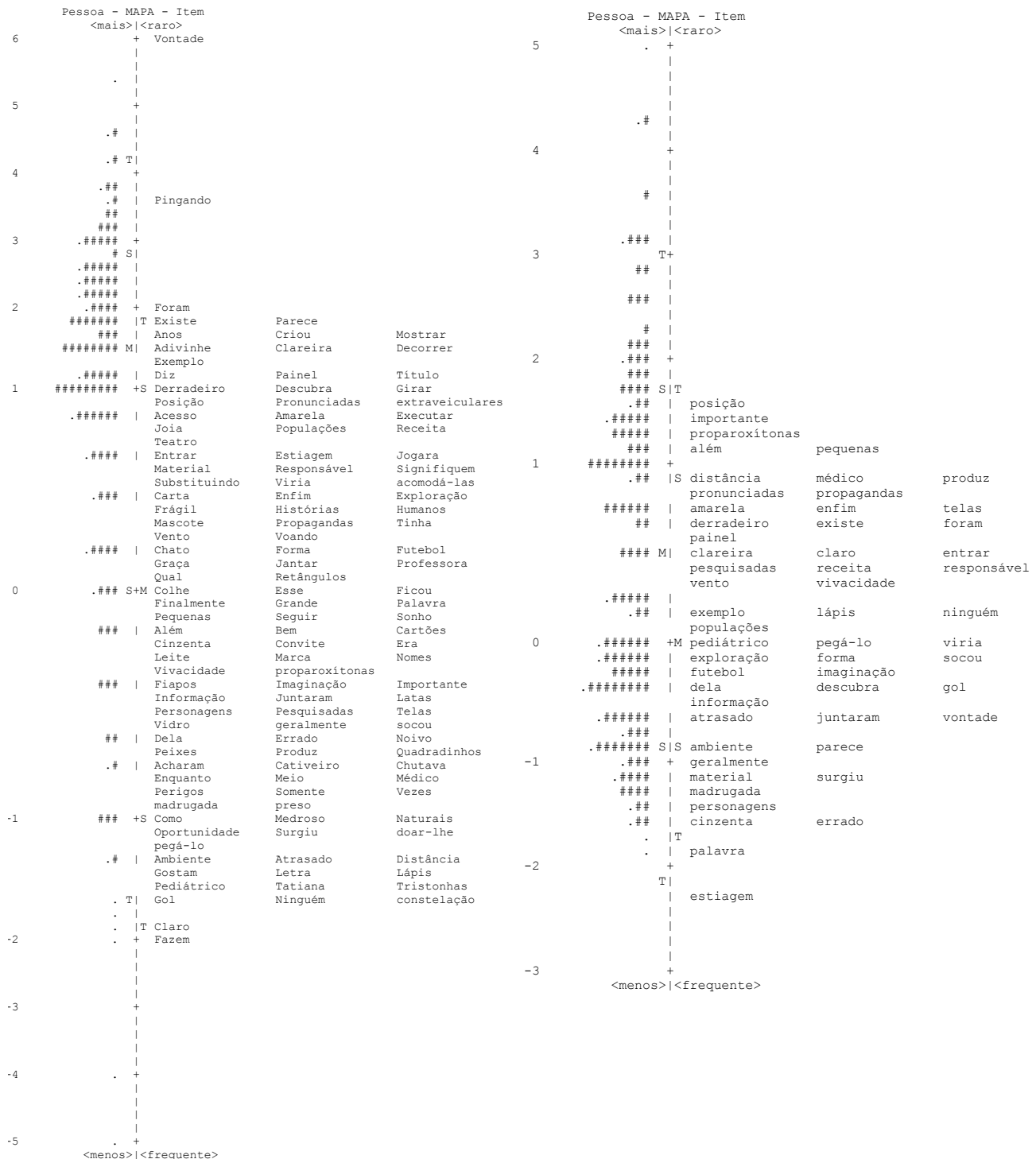


Figura 1. Mapa das crianças e itens na escala de 126 e 55 itens

Tabela 4

Exemplos de itens que foram mantidos ou eliminados em razão de suas dificuldades e ajustes

Itens	Dificuldade	Infit	Outfit	Itens	Dificuldade	Infit	Outfit
Receita	0,87	0,93	0,86	Descubra	1,04	1,06	1,19
Teatro	0,87	1,44	1,54	Extraveiculares	1,01	1,52	1,95
Amarela	0,75	0,94	0,85	Clareira	1,37	1,31	1,49
Acesso	0,75	1,23	1,27	Adivinhe	1,35	1,14	1,30
Jogara	0,65	1,20	1,25	Decorrer	1,34	1,07	1,05
Material	0,65	1,07	1,06	Exemplo	1,34	1,10	1,16

Discussão

O objetivo deste estudo foi dar continuidade ao desenvolvimento de uma escala de reconhecimento ortográfico de palavras, mediante a validação cruzada. Diversos itens foram eliminados do instrumento com base nos critérios de índice de dificuldade, ajuste e tipos de dificuldades linguísticas contempladas nas opções de resposta. Após as mudanças no instrumento, procedeu-se a uma nova coleta de dados, com vistas a averiguar algumas propriedades psicométricas da nova versão.

No contexto da aquisição da leitura, as habilidades de leitura de palavras, compreensão linguística e de reconhecimento ortográfico de palavras escritas devem ser facilitadas como consequência do processo de alfabetização e consolidação do conhecimento linguístico (Berninger et al., 2002); por isso, espera-se que as palavras do instrumento sejam mais fáceis para os alunos de anos mais avançados que de anos anteriores. No entanto, análises do instrumento inicial de 221 itens mostraram que esse resultado esperado não foi obtido empiricamente para alguns itens.

A presença de palavras com esse problema torna o instrumento menos eficaz e menos condizente com o construto que se deseja avaliar, uma vez que a aquisição do conhecimento ortográfico é gradual e a escala de reconhecimento ortográfico deve ser sensível às mudanças consequentes do tempo de escolarização. Então, mesmo que as crianças estejam em um momento de instabilidade e acertem e errem palavras inesperadas, os resultados mostraram que há consistência na resposta dos estudantes em uma série de palavras. Manter essas palavras no instrumento favorece a sua qualidade. Por isso, optou-se pela eliminação de palavras que são respondidas de forma inconsistente pelas crianças.

Após a eliminação das palavras com índices de dificuldade problemáticos, os itens restantes foram submetidos a uma análise pelo modelo de Rasch. Observou-se que o instrumento como um todo apresentou um bom ajuste ao modelo de Rasch, indicando que as crianças responderam aos itens de acordo com o esperado, ou seja, crianças com alto nível de habilidade acertaram itens mais difíceis, enquanto crianças com menos habilidades acertaram itens mais fáceis. No entanto, dois

itens apresentaram desajuste moderado em relação ao *infit* e seis itens desajustaram no que diz respeito ao *outfit*. Quatro deles foram desajustes moderados e dois foram altos. Os desajustes revelados por ambos os indicadores mostram que esses itens receberam respostas inesperadas por parte das crianças, o que quer dizer, que, por exemplo, crianças com nível de habilidade mais baixo acertaram itens mais difíceis (Linacre, 2002). A quantidade de itens desajustados não foi preocupante e pode-se entender que a eliminação dos itens que obtiveram respostas inconsistentes, medidas pelos índices de dificuldade, foi eficaz, pois gerou um novo conjunto de itens que, em média, se ajustou às respostas esperadas.

Não obstante, o processo de desenvolvimento do instrumento ainda passou por mais uma etapa, em que foram realizadas análises dos parâmetros de dificuldade das palavras pelo modelo Rasch. A eliminação dos itens com base na dificuldade não reduziu o intervalo de construto representado por eles, mas eliminou-se aqueles que coincidiam na localização. Essa etapa do processo de revisão dos itens foi criteriosa e levou em consideração sua redundância, além do ajuste e da dificuldade linguística representada na palavra, de acordo com os estudos de Sisto (2001) e Caliatto (2005), que desenvolveram categorias de erros ortográficos da língua portuguesa. Além disso, houve uma preocupação por manter a maior variedade possível das dificuldades encontradas pelos alunos na aprendizagem da língua portuguesa, considerando, no entanto, a limitação do formato do instrumento e do número de palavras.

Com o intuito de averiguar se essa etapa do processo gerou um conjunto de itens que se adequem a uma escala, foi realizada nova análise de Rasch. Os parâmetros de ajuste indicaram que houve uma melhora considerável do ajuste dos itens ao modelo, já que uma quantidade muito menor de itens apresentou desajuste. Novamente, isso indica que os itens foram respondidos conforme o previsto pelo modelo.

Ainda, a comparação visual entre os 2 mapas do construto mostra que a escala de 55 itens manteve um amplo intervalo da escala representado por itens com diferentes níveis de dificuldade. Ressalta-se, no entanto, que o mapa indica que há poucos itens com dificuldades elevadas, o que pode produzir avaliações pouco

precisas de crianças com habilidades ortográficas altas. Observa-se, ainda, que não há acúmulo de itens localizados no mesmo intervalo de dificuldade, o que evita que as crianças respondam a itens desnecessários.

No que diz respeito à capacidade da escala com 55 itens para diferenciar o desempenho dos alunos dos diferentes anos do ensino fundamental, foi realizada uma ANOVA com os parâmetros de habilidades das crianças em função dos anos escolares. Efetivamente, a análise mostrou que a escala revelou diferenças significativas entre os alunos de todos os anos escolares, o que foi interpretado como evidência de validade de critério da Escala de Avaliação de Reconhecimento de Palavras – Versão computadorizada de 55 itens. Em relação ao processo de validação cruzada, esse resultado indicou que o conjunto de itens selecionado foi mais sensível à aquisição ortográfica dos alunos em comparação com o instrumento inicial.

Finalmente, a nova versão de 55 itens da escala está em consonância com as considerações teóricas sobre a aquisição das habilidades escolares e o desenvolvimento das habilidades cognitivas propostas por Berninger et al. (2002). Por exemplo, espera-se que crianças de quarto e quinto anos estejam aprendendo e desenvolvendo as estratégias contextualizadas de escrita de forma mais eficaz que as crianças mais novas. Igualmente, entende-se que os alunos mais velhos devem fazer melhor uso de sua memória de trabalho que crianças mais novas, devido à estimulação das habilidades cognitivas que o processo de escolarização proporciona e à própria idade, ligeiramente mais avançada dos alunos (Urquijo, 2010). Ao mesmo tempo em que essas habilidades são desenvolvidas, elas auxiliam a aquisição dos processos de escrita, como a geração de texto e a transcrição, entendida como integração da escrita motora com a ortografia (Berninger et al., 2002).

Como considerações finais desta pesquisa, pode-se afirmar que o desenvolvimento da nova versão foi eficaz em relação ao ajuste dos itens e à diferenciação do desempenho dos alunos em função dos anos escolares. Isso possibilita compreender a Escala de Avaliação de Reconhecimento de Palavras inserida no contexto de um modelo de aquisição gradual do conhecimento da língua, conforme proposto por Berninger et al. (2002). Entende-se que a escala avalia o conhecimento ortográfico das crianças em fase de alfabetização, de forma isolada da ação motora da escrita. Assim, a utilidade desse tipo de avaliação se deve à possibilidade de mensurar apenas o conhecimento da ortografia que a criança possui, sem a necessidade da escrita, que é muito laboriosa nos anos iniciais do ensino fundamental. Dessa forma, o instrumento pode ser considerado uma alternativa ao ditado, quando o interesse da avaliação é a ortografia e não a grafia dos alunos. Assim, a escala pode auxiliar na avaliação de crianças em fase de alfabetização em conjunto com outros tipos de estratégias de avaliação.

No que diz respeito à limitação do estudo, não foram calculados os valores de discriminação dos itens. Como perspectivas futuras, mais estudos são necessários para investigar outras propriedades psicométricas da nova versão da escala, como precisão, dimensionalidade, entre outros, e buscar diferentes tipos de evidências de validade. É interessante também considerar outros modelos da Teoria de Resposta ao Item (TRI) para análise da escala, como, por exemplo, o modelo de dois parâmetros, já que contempla índices diferenciados de discriminação para os itens. Finalmente, entende-se que são necessárias novas coletas para obtenção de dados em amostras com diferentes características e observância da permanência das propriedades psicométricas e vantagens do uso da Escala de Avaliação de Reconhecimento de Palavras.

Referências

- Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1-29. doi: 10.1146/annurev-psych-120710-100422
- Báez, A. L. M. (2013). Cuestionario de estrategias para la escritura de ensayos argumentativos. *Acta Colombiana de Psicología*, *16*(1), 83-98. Recuperado de <http://www.scielo.org.co/pdf/acp/v16n1/v16n1a13.pdf>
- Berninger, V. W., Abbott, R. D., Abbott, S. P., Graham, S., & Richards, T. (2002). Writing and reading: Connections between language by hand and language by eye. *Journal of Learning Disabilities*, *35*(1), 39-56. doi: 10.1177/002221940203500104
- Berninger, V. W., Richards, T. L., & Abbott, R. (2015). Differential diagnosis of dysgraphia, dyslexia, and OWL LD: Behavioral and neuroimaging evidence. *Reading and Writing*, *28*(8), 1119-1153. doi: 10.1007/s11145-015-9565-0
- Brooks, A. D., Berninger, V. W., & Abbott, R. D. (2011). Letter naming and letter writing reversals in children with dyslexia: Momentary inefficiency in the phonological and orthographic loops of working memory. *Developmental Neuropsychology*, *36*(7), 847-868. doi: 10.1080/87565641.2011.606401
- Cagliari, L. C. (2010). *Alfabetização e linguística*. São Paulo: Editora Scipione.
- Cain, K., Bryant, P., & Oakhill, J. (2004). Children's reading and comprehension ability: Concurrent prediction by Working Memory, Verbal Ability, and Component Skills. *Journal of Educational Psychology*, *96*(1), 31-42.
- Caliatto, S. G. (2005). *Avaliação da escrita em jovens e adultos* (Dissertação de mestrado, Universidade Estadual de Campinas, Campinas, SP, Brasil). Recuperado de <http://www.bibliotecadigital.unicamp.br/document/?code=vtls000375590>

- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135-168. doi: 10.1146/annurev-psych-113011-143750
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum Associates.
- Fernandes, D. C., Sisto, F. F., & Murarolli, P. L. (2014). Escala de avaliação do reconhecimento de palavras: versão computadorizada – versão computadorizada (programa de computador). Pouso Alegre: Univás.
- Jones, J. N., J., Abbott, R. D., & Berninger, V. W. (2014). Predicting levels of reading and writing achievement in typically developing, english-speaking 2nd and 5th graders. *Learning and Individual Differences*, 32, 54-68. doi: 10.1016/j.lindif.2014.03.013
- Kingski, M. F., & Sisto, F. F. (2004). Reconhecimento de palavras nas séries iniciais do ensino fundamental. *Revista Teoria e Prática da Educação*, 7(2), 173-184.
- Kintsch, W., & Dijk, T. A. V. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363-394.
- Linacre J. M. (2002). What do Infit and Outfit, Mean-Squared and Standardized mean? *Rasch Measurement Transactions*, 16(2), 878. Recuperado de <http://209.238.26.90/rmt/rmt82a>
- Linacre, J. M. (2009). *A user's guide to Winsteps & Ministeps*. Rasch Model Computers Programs. Chicago: MESA Press.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedagogiske Institut.
- Sisto, F. F. (2001). Dificuldades na aprendizagem em escrita: um instrumento de avaliação (ADAPE). Em F. F. Sisto, E. Boruchovitch, L. D. T. Fini, R. P. Brenelli, & S. C. Martinelli (Orgs.), *Dificuldades de aprendizagem no contexto psicopedagógico*. (pp. 190-213). Petrópolis: Vozes.
- Urbina, S. (2007). *Fundamentos da testagem psicológica*. Porto Alegre: ArtMed.
- Urquijo, S. (2010). Funcionamento cognitivo e habilidades metalinguísticas na aprendizagem. *Educar em Revista*, 38, 19-42.

recebido em dezembro de 2015
reformulado em fevereiro de 2016
aprovado em abril de 2016

Sobre os autores

Débora Cecílio Fernandes é graduada em Psicologia pela Universidade Estadual de Londrina, Doutora em Neuropsicologia Clínica pela Universidad de Salamanca, Espanha. Atualmente, é professora da Universidade Estadual do Centro-Oeste (Unicentro), Guarapuava, Paraná. Participa do Grupo de Estudos e Pesquisa em Psicopedagogia da Universidade Estadual de Campinas (Unicamp).

Fermino Fernandes Sisto é graduado em Pedagogia, Doutor em Psicologia pela Universidad Complutense de Madrid. Professor aposentado Titular da Universidade Estadual de Campinas (Unicamp).