

ESTUDO DO FUNCIONAMENTO DIFERENCIAL DE ITENS PARA AVALIAR O RECONHECIMENTO DE PALAVRAS

Fermino Fernandes Sisto¹ – Universidade São Francisco

RESUMO

Para avaliar a presença de funcionamento diferencial em função da variável sexo, foram analisados 221 itens, construídos para mensurar o nível de reconhecimento de palavras. O estudo foi realizado com 332 crianças, com idades entre 7 e 11 anos, freqüentando de primeira a quarta série do ensino fundamental dos sistemas público e particular de educação. O instrumento possuía três alternativas de resposta, das quais somente uma era a correta, para a qual foi atribuído um ponto. Os índices de precisão do instrumento foram de 0,98 para os itens e 0,97 para as pessoas, e alfa de Cronbach de 0,98. Apenas 4 itens apresentaram DIF, sendo que pelo método de Mantel-Haenszel apenas três itens, todos eles apontados pelo procedimento anterior. No geral, menos de 2% do total de itens do instrumento apresentou funcionamento diferencial de itens.

Palavras-chave: reconhecimento de palavra; funcionamento diferencial do item; viés; avaliação.

STUDY OF THE DIFFERENTIAL FUNCTIONING OF ITEMS TO ASSESS WORD RECOGNITION

ABSTRACT

To evaluate the presence of differential functioning as function of the sex variable, 221 items were prepared to assess the level of recognition of words and analyzed. The study was accomplished with 332 children, with ages between seven to 11 years, attending at first-fourth grades of basic education of public and private schools. The test had three alternatives for responding, and for the correct response one point was attributed. The reliability coefficients were .98 to the items and .97 to the persons, and the alpha of Cronbach was .98. The first analysis indicated only four items with differential item functioning, but the Mantel-Haenszel technique pointed out three items, all of them indicated by the previous procedure. In general, less than 2% of the total of the items of the test presented differential functioning.

Keywords: Word recognition, differential item functioning, bias, assessment.

INTRODUÇÃO

O reconhecimento de palavras tem sido um tema de pesquisa que, por décadas, desperta o interesse de muitos estudiosos. Sua importância pode ser constatada na vasta gama de estudos que relacionam o reconhecimento de palavras com outras variáveis, tais como retardo mental (Calhoun, 1985), mal de Parkinson (Garcia de la Cadena, Ostrosky Solis, Rodríguez, Madrazo & outros, 1992), Alzheimer (Motomura, Tomota, Akagi & Seo, 2000) e dislexia (Gallagher, Laxon, Armstrong & Frith, 1996).

O reconhecimento de palavras, habilidades fonológicas e ortográficas são variáveis comuns a muitos estudos e formam um grupo importante de pesquisas. O papel da consciência fonológica na aquisição da escrita e da leitura está em estreita relação com reconhecimento de palavras e soletração. Dentre os principais resultados, pode-se dizer que, no geral, sujeitos com dificuldades em reconhecimento de letras apresentaram dificuldades

em apagar e contar fonemas; leitores pouco habilidosos apresentaram baixa velocidade de identificação e pouca consciência fonológica e necessitavam de mais tempo para avaliarem os pares de palavras com incongruências fonológicas; a velocidade de identificação de letras está relacionada com o reconhecimento de palavras; e similaridades de formas e sons de palavras de mesmo radical interagem no reconhecimento de seus caracteres (Assink & Kattenberg, 1995; Badian, 1993; Lukatela, Carello, Shankweiler & Liberman, 1995; Mahony, Singson & Mann, 2000).

O reconhecimento de palavras estaria mais relacionado com a variável soletração para os alunos dos níveis elementares e aprendizes em leitura (Shanahan, 1984); enquanto que leitores mais jovens e menos habilidosos seriam mais sensíveis aos princípios da análise silábica em reconhecimento de palavras do que aos princípios morfológicos (Leong & Parkinson, 1992); como também o reconhecimento de palavras na leitura letra-a-letra seria mediado pelo léxico ortográfico de entrada utilizado na leitura normal (Rapcsak, Rubens & Laguna, 1990).

¹ Endereço para contato: E-mail: fermino.sisto@gmail.com

Além da soletração, a influência da extensão da palavra em sua identificação tem sido estudada. A correlação entre o tamanho da unidade ortográfica e os mecanismos de processamento que distinguem os leitores habilidosos dos pouco habilidosos, Gottardo, Chiappe, Siegel e Stanovich (1999) apontaram que os problemas relativos à codificação soletração-som, podiam ser interpretados como a causa das dificuldades em leitura.

Em seus experimentos sobre o reconhecimento de palavras, Hung e Tzeng (1981) se interessaram em investigar como os grafemas eram mapeados pela linguagem falada. De acordo com seus resultados, o processamento da informação visual era realmente afetado pela variação ortográfica, mas só nos níveis mais baixos. Com relação ao processamento de alto nível, o comportamento de leitura foi imune às variações ortográficas. Por sua vez, Rey, Ziegler e Jacobs (2000) constataram que os grafemas eram processados pelo sistema de leitura do indivíduo como unidades perceptuais. Nesse contexto, ao realizar uma pesquisa para investigar se as palavras eram reconhecidas como um todo ou por partes, Gough (1993) indicou que as primeiras palavras eram reconhecidas por associação seletiva, pois foram reconhecidas as metades de cada palavra.

A relação fonema-grafema também vem sendo estudada no Brasil, porém as pesquisas realizadas ainda são poucas, recentes e sua maior parte relacionada com a alfabetização e não especificamente com o reconhecimento de palavras. Há trabalhos como o de Carvallo, Martes, Camargo, Chiari e Vieira (1988) que averiguaram os padrões sequenciais das letras da língua portuguesa escrita; a relação entre consciência fonológica e aquisição da linguagem escrita (Santos, 2002) e a aquisição do princípio alfabético na escrita (Mota, Moussatche, Castro, Moura & D'Angelis, 2000).

As pesquisas em disléxicos reforçam os achados das pesquisas de outros países. Assim, entre outros, Kajihara (1997) afirmou que os indivíduos portadores da dislexia apresentavam atraso no desenvolvimento da escrita, da consciência fonológica e da memória verbal, prejuízo no uso da via sublexical de leitura e utilizavam códigos fonológicos para a memória de curto prazo; Pestun (2001) também encontrou prejuízos lingüísticos, cognitivos e várias outras deficiências, incluindo habilidades de leitura e escrita; e Pinheiro (1995) sugeriu que a dificuldade principal dos indivíduos disléxicos parecia estar

associada a déficits na memória verbal e nos processos de segmentação e síntese de fonemas.

Com relação aos aspectos da aprendizagem em reconhecimento de palavras, Góes (1984) observou níveis progressivos da criança, em fase de alfabetização, para lidar com as correspondências som-texto. Por sua vez, Graminha, Machado, Francischini e Befi (1987) também se preocuparam com a eficiência das crianças em reconhecer palavras. Demonstraram que é possível melhorar o desempenho de crianças, em fase de alfabetização, com dificuldades específicas na discriminação de sílabas semelhantes, tanto na leitura quanto na escrita.

Com relação à avaliação do reconhecimento de palavras e suas características por meio de instrumentos, Jaeger, Schosler e Wainer (1998) estudaram a aquisição da escrita em crianças e em adultos, e constataram que as crianças também tiveram um melhor desempenho, inclusive sob a ótica da familiaridade semântica dos vocábulos, tanto para palavras simples quanto complexas.

Em relação aos estudos que abordam a problemática fonema-grafema, direta ou indiretamente, podem ser mencionados os de Galdo (1989) que elaborou uma bateria de testes para diagnóstico (BTDP) para determinar níveis de habilidades básicas necessárias para a aquisição da leitura e da escrita em pré-escolares; Oliveira (1996) que constatou a fidedignidade e a validade preditiva e concorrente de três testes distintos, quais sejam, o Teste de Diagnóstico de Habilidades do Pré-Escolar, o Teste de Prontidão para Leitura e o Instrumento para Avaliação do Repertório Básico para Alfabetização; Sisto (2001) que construiu um instrumento para avaliar os níveis de dificuldades a escrita de crianças em processo de alfabetização, denominado ADAPE (Avaliação de Dificuldades na Aprendizagem da Escrita), circunscrito às dificuldades de representação de fonemas, ou seja, grafar letras e palavras com base em um sistema lingüístico estruturado e que apresenta arbitrariedades; e Pinheiro e Rothe-Neves (2001) desenvolveram o Procedimento de Avaliação Cognitiva da Leitura e da Escrita, que se revelou adequado para a avaliação da leitura e da escrita em crianças da primeira à quarta série do ensino fundamental.

Em face de esse quadro da realidade brasileira, pode-se afirmar que embora existam vários instrumentos usados para avaliar o reconhecimento de palavras nas diferentes pesquisas, não se localizou estudos voltados para

identificar as evidências de validade dos instrumentos usados. De fato, há uma carência de pesquisas na área de reconhecimento de palavras, pese a sua importância não apenas na área educacional, quando se estuda o processo de alfabetização e as dificuldades para sua aquisição, problema hoje de grande vulto na nossa realidade de ensino público, mas também como auxiliar no diagnóstico de dislexia, mal de Parkinson, Alzheimer, entre outros.

A literatura estrangeira também registra uma variedade grande de instrumentos nas pesquisas, criados para analisar aspectos específicos que interessam ao pesquisador estudar e a maior parte das pesquisas usam esse procedimento. Em contraposição, os estudos de validade dos instrumentos para avaliar reconhecimento de palavras não são muitos.

Alguns exemplos dos poucos estudos de validação de instrumentos dessa natureza são comentados a seguir e indicam a direção das pesquisas. Freyberg (1970), utilizando como critério o desempenho em soletração, validou dois testes, um de palavras ditadas (relacionado à recordação) e um teste de múltipla escolha (relacionado ao reconhecimento). Segundo seus resultados, o teste de palavras ditadas obteve uma validade simultânea mais alta do que o de múltipla escolha. Por sua vez, com o objetivo de padronizar medidas para habilidades de reconhecimento de palavras em crianças, Cunningham e Stanovich (1993) encontraram que as habilidades de processamento ortográfico e fonológico deviam ser consideradas como componentes distintos no reconhecimento de palavras. Entretanto, nesse contexto, Moss, Cole e Khampalikit (1982) sugeriram a necessidade de limites claros na generalização dos resultados entre as formas de avaliações de habilidades de escrita para casos de indivíduos de diferentes séries. Já Flanagan e cols. (1997) investigaram os testes Kaufman Functional Academic Skills Test (K-FAST), Woodcock-McGrew-Werder Mini-Battery of Achievement (MBA) e Wide Range Achievement Test-3 (WRAT-3). Os resultados forneceram a evidência de validade simultânea para os três testes, sendo que o MBA proporcionou uma avaliação mais ampla para a leitura (reconhecimento de palavras, compreensão da escrita e compreensão de passagem) e a escrita (pontuação, capitalização, soletração e utilização da palavra), enquanto o WRAT-3 avaliou principalmente o reconhecimento

de palavras, apresentando uma alta correlação com o MBA.

A respeito da questão relacionada aos estudos de validade, Messick (1980, 1989) realçou que tipos distintos de inferências com os testes demandam diferentes tipos de evidência, nada obstante tenha defendido um conceito unitário de validade. Assim, as evidências poderiam ser obtidos investigando as relações entre as respostas e as tarefas, examinando o conteúdo do teste em função dos conteúdos de referência, estudando as relações entre as pontuações do teste e outras medidas ou temporalmente, investigando as diferenças entre os grupos ou as respostas dos sujeitos a tratamentos experimentais, por exemplo.

Também, é necessário considerar que as posições atuais (AERA, APA & NCME, 1999) sobre a validade dos testes destacam principalmente que o que se valida são as pontuações e não o teste, ou seja, a pergunta que se faz quando se analisa seria se o uso ou a interpretação das pontuações de um teste são válidos. Nesse sentido, a teoria tem um papel muito importante como orientação tanto o desenvolvimento de um teste como de seu processo de validação. Ao lado disso, a validação é um processo contínuo e dinâmico; assim, a evidência de validade de um teste não pode se resumir apenas a um indicador ou índice, mas ela é assegurada pelo acúmulo teórico, estatístico e empírico do uso de suas pontuações. Também hoje é bastante aceito que uma pontuação pode ser válida para uma determinada situação sem o ser para outra.

Nesse contexto é importante exigir dos instrumentos, sejam eles projetivos, de execução, questionários, por exemplo, informações sobre sua validade, sem se esquecer também de sua precisão. A validade é algo de muita importância não apenas para ser analisada depois que o instrumento esteja pronto, mas também durante seu processo de construção, pois é ela que permitirá ou que dará respaldo às inferências e interpretações a serem feitas com base nas pontuações que obtidas com a aplicação de um instrumento, já as pontuações consubstanciam a relação entre o constructo/variável que se pretende medir.

Em razão principalmente de questões sociais, nos últimos tempos tem sido gerado uma extensa investigação, tendo como foco o fato de que muitos testes foram acusados de estarem enviesados. Isso significaria que, em igualdade de condições, as pessoas pertencentes a um grupo (caso se considere a variável sexo, por exemplo, meninos) obteriam pontuações

sistematicamente menores ou maiores no instrumento em questão que as pessoas do outro grupo (meninas). Tais estudos, popularmente conhecidos como estudos de vieses, já algum tempo se prefere denominá-lo de *funcionamento diferencial do item* (DIF) (Holland & Thayer, 1988).

É considerado que um item apresenta DIF ou funciona diferentemente quando dois grupos de pessoas, semelhantes em relação a um atributo (por exemplo, reconhecimento de palavras) não o acertem na mesma proporção. Mellenbergh (1982, 1989) distinguiu dois diferentes tipos de DIF, sendo que é denominado *uniforme ou consistente* quando não existe interação entre o nível da habilidade medida e o fato de pertencer a um grupo; e outro é denominado de DIF *não uniforme* ou *inconsistente* se ocorrer interação, ou seja, a probabilidades de responder corretamente a um item nos dois grupos não é a mesma em todos os níveis da habilidade.

Nesse contexto, a investigação busca determinar se algum tipo de característica (nível socioeconômico, sexo, raça, entre outros), reconhecidamente extemporânea ao funcionamento cognitivo e psicológico, tem um efeito sobre a medida de um traço psicológico. Assim, instrumentos de medida que desfrutam de um papel proeminente no diagnóstico de conduta disfuncional ou seleção de pessoal, demandam a identificação e análise de seus itens para compreender como as características demográficas das pessoas e bagagem sociocultural interagem no processo de medida de variáveis psicológicas.

Ao mesmo tempo, é necessário considerar que, em sua função de medir, um instrumento não deveria ser afetado pelas características do objeto a ser medido. Dependendo do grau em que esteja contaminado, sua validade estará gravemente comprometida. Isso é um sério problema, pois a objetividade da medida é um requisito imprescindível para qualquer teste ou escala psicológica.

Estabelecer a validade de um teste implica também em obter evidência de que o instrumento com o qual se trabalha está livre de vieses. Em outros termos, os itens do teste devem funcionar de igual maneira para diferentes grupos independentemente de variáveis que possam se constituir em uma fonte sistemática de variação, alheia ao constructo medido pelo teste.

Em relação à problemática de escrita e leitura, envolvendo ou não o reconhecimento de palavras, a literatura tem indicado sistematicamente que os meninos apresentam mais dificuldades de aprendizagem que as meninas (Bartholomeu, Sisto & Rueda, 2006; Bazi & Sisto, 2006; Kingenski & Sisto, 2004; Rueda, Bartholomeu & Sisto, 2004; Sisto, F. F., 2001; Sisto, Boruchovitch, Brenelli, Fini, Martinelli & Urquijo, 2001; Sisto, & Fernandes, 2004). A pergunta que se coloca é se os itens usados estariam livres de vieses e realmente os meninos apresentam maiores dificuldades para lidar com grafema-fonema, ou se os itens estariam enviesados e por isso um grupo estaria sendo privilegiado em sua avaliação.

Assim, o objetivo deste trabalho foi avaliar a possível presença de funcionamento diferencial em um conjunto de itens construídos para mensurar o nível de reconhecimento de palavras em crianças em fase de escolarização, em função da variável sexo. Trata-se de um estudo inicial, pois a escala está em seu início de construção e se pretende que apenas a habilidade de reconhecimento prepondere ao longo dos itens e não a variável sexo, por considerar que o funcionamento diferencial de alguns itens podem comprometer sua interpretação, como também porque se considera a análise de DIF como uma fase do processo de elaboração de instrumentos.

MÉTODO

Participantes

O estudo foi realizado com 332 crianças, ambos os sexos (56,8% de meninas), do sistema público e particular de educação, em uma cidade do interior paulista. As idades variaram de 7 a 11 anos, com uma média de 8,81 (desvio padrão de 1,33) e moda e mediana de 9 anos; sendo que os percentuais por idade variaram entre 10,40% (11 anos) a 24% (9 anos). Frequentavam de primeira a quarta série do ensino fundamental, tendo todos eles cursado a pré-escola e iniciado a alfabetização na primeira série. A distribuição das crianças por série foi bastante equilibrada, distribuída no intervalo de 22,80 a 27,20%.

Instrumento

Para a construção do instrumento foram sorteadas, equiprovável e aleatoriamente, 20% das páginas dos materiais fornecidos por escolas. Das

palavras dessas páginas, foram sorteadas, seguindo o mesmo procedimento, 20% delas, perfazendo um total de 1005 palavras sorteadas. Valendo-se desse conjunto, as palavras foram distribuídas, sempre com o mesmo procedimento, em quatro subconjuntos, sendo que um deles foi o escolhido por sorteio para compor o instrumento de avaliação para o reconhecimento de palavras.

Foram excluídas 17 palavras por semelhanças com outras palavras (por exemplo, *letra e letras*) mediante sorteio, restando um total de 221 palavras. Com base nos critérios de classificação de Sisto (2001), verificou-se que 56 palavras não possuíam qualquer tipo de dificuldade de escrita (25,34%) ao passo que 165 delas possuíam dificuldades de escrita (74,66%).

Com relação à formatação do instrumento, as palavras foram organizadas em ordem alfabética. Para cada palavra foram criadas duas outras com erros gramaticais. Assim, o instrumento tomou a forma de um teste de múltipla escolha, composto por 221 itens, sendo que cada item possuía três alternativas de resposta, das quais somente uma era a correta. Para cada acerto foi atribuído um ponto, sendo possível atingir o escore máximo de 221 pontos.

A aplicação foi coletiva, após ter conseguido o termos de consentimento dos pais, sendo que o professor da sala lia a palavra e, em seguida, a criança deveria marcar com um risco a grafia correta da palavra no instrumento fornecido para elas. Em cada classe houve dois auxiliares de aplicação, bolsistas do curso de psicologia. Antes de iniciar a aplicação propriamente dita, foi fornecido um exemplo para explicação para os alunos.

Análise dos dados

Para estudar o funcionamento diferencial dos itens os dados foram analisados por dois modelos, ainda que exista um grande número deles (Fidalgo, 1996). Um deles foi desenvolvido por Mantel e Haenszel (1959) e tem por base a comparação de frequências observadas e esperadas nos grupos em estudo nos vários subgrupos constituídos por pessoas com habilidades semelhantes. Sua avaliação é feita pelo nível de significância das diferenças. O outro, estima quanto de DIF é acrescido (positiva e negativamente) ao item e, por meio da prova *t* de Student, calcula se a diferença pode ser atribuída ao acaso ou não. Em geral, um valor de *t* maior do que dois é considerado significativo. No entanto, Draba (1977) considera, concordando com Bonferroni, que o patamar de 2,4 é um ponto bom para a significância estatística, quando se está analisando mais de 20 itens. O programa estatístico usado para fazer ambas as análises foi o Winsteps.

RESULTADOS

Após a aplicação e correção dos instrumentos, os dados foram analisados pelo modelo Rasch, com vistas a ver sua adequação a uma escala. Os resultados em relação à precisão do instrumento forneceram um índice de 0,98 para os itens e 0,97 para as pessoas, como também um alfa de Cronbach de 0,98. Esses valores indicam uma precisão ótima para ambos os casos. Por sua vez, o erro médio da medida das pessoas foi de 0,20 (DP=0,07) e dos itens foi de 0,16 (DP=0,03). Mais detalhes podem ser observados na Tabela 1.

Tabela 1. Parâmetros de ajuste dos itens e das pessoas

| Parâmetros | Itens | | | Pessoas | | |
|-------------|--------------|---------------|-------------|--------------|---------------|-------------|
| | <i>Infit</i> | <i>Outfit</i> | <i>erro</i> | <i>Infit</i> | <i>Outfit</i> | <i>erro</i> |
| Media | 0,99 | 0,94 | 0,16 | 0,98 | 0,94 | 0,20 |
| D.P. | 0,16 | 0,45 | 0,03 | 0,12 | 0,34 | 0,07 |
| Máximo | 1,62 | 5,48 | 0,25 | 1,43 | 6,8 | 0,72 |
| Mínimo | 0,72 | 0,25 | 0,12 | 0,74 | 0,27 | 0,15 |
| 1,3 < > 1,5 | 10 (3,01%) | 12 (3,61%) | | 8 (2,41%) | 14 (4,22%) | |
| 1,5 < > 2,0 | 1 (0,03%) | 12 (3,61%) | | 0 (0,00%) | 21 (6,33%) | |
| < 2,0 | 0 (0,00%) | 3 (0,90%) | | 0 (0,00%) | 3 (0,90%) | |

No que tange aos itens, a média do *infit* foi 0,99 (DP=0,16), indicativo que no geral a maioria deles foi respondida de acordo com o padrão esperado que seria de 1,00. Em uma análise mais detalhada, observou-se que os valores *infit* ficaram

entre 0,72-1,62 deixando transparecer que nem todos os itens se enquadraram no intervalo 0,70-1,30 que normalmente é considerado bom ajuste. Em uma análise dos itens que mostraram desajuste, observou-se que 11 deles forneceram um *infit*

apenas 4,51% teriam apresentado discrepâncias preocupantes.

No que concerne às pessoas, a maioria das pessoas apresentou o padrão esperado que seria de 1,00, pois a média do *infit* foi 0,98 (DP=0,12) e a do *outfit* de 0,94 (DP=0,34). Entretanto, o intervalo de variação do *infit* (0,74-1,43) foi bastante diferente da variação do *outfit* (0,27-6,8). Em relação ao *infit*, pode-se considerar que nenhuma criança apresentou desajuste pelos critérios de Linacre (2002), diferentemente dos resultados do *outfit*, cujos dados informaram que 24 crianças (7,23%) mostraram padrões não ajustados à escala. De forma geral, os percentuais de desajuste podem ser considerados pequenos, principalmente por se tratar de um estudo de uma escala em construção.

Em relação ao intervalo de distribuição dos itens na escala, houve uma variação de -1,96 a 3,66. Apesar de esse resultado indicar itens muito fáceis e muito difíceis, a extensão de 5,62 pode ser considerada pequena para a quantidade de itens analisada. Entretanto, como pode ser observado na Tabela 2, a maioria deles se agrupou entre os valores +1 e -1 da escala *logit*, intervalo esse que avalia as pessoas com maior precisão e menor erro.

Ao lado disso, a média da habilidade das crianças foi de 1,47 (DP=1,37) e dos itens 0,00 (DP=1,02), sugerindo que o instrumento foi relativamente fácil para essas pessoas. Esse dado

pode ser visualizado na Tabela 2, onde a letra M, ao lado da divisória das informações das pessoas e dos itens, indica a localização da média.

Como proposto, os itens foram analisados por dois métodos. Na Tabela 3 estão resumidos os resultados que apresentaram significância estatística em um ou outro método ou em ambos. Do total de 211 itens, apenas 10 deles forneceram resultados significativos, o que corresponde a 4,7%, proporção essa considerada pequena em uma quantidade de itens como foi a estudada. Nesse sentido, a variável sexo parece estar influenciando em poucos itens o reconhecimento de palavras. Analisando pelo critério nível de significância, o método proposto por Linacre no Winsteps forneceu 10 itens com possibilidade de viés. No entanto, usando o critério de Dabra (1977) para esse procedimento, apenas os itens com *t* superior a 2,40 deveriam ser considerados problemáticos. Assim, teríamos que considerar que apenas 4 itens estariam nessa condição. Pelo procedimento de Mantel-Haenszel apenas três itens apresentaram diferenças estatisticamente significativas e os três foram apontados pelo procedimento anterior. Esses dados parecem indicar que entre três e quatro itens poderão ter um viés na medida do reconhecimento de palavras, o que corresponderia a menos de 2% do total de itens do instrumento.

Tabela 3. Medidas de DIF por sexo, mudança e valores de *t* e probabilidade e tamanho pelo método Mantel-Haenszel, por item.

| sexo | DIF medida | | mudança | <i>t</i> * | Prob. | Prob** | tamanho | Item |
|-------|------------|------------|---------|------------|--------|--------|---------|------|
| | Rasch | sexo Rasch | | | | | | |
| Masc. | 1,78 | Fem. 3,16 | 1,38 | 5,17 | 0,0000 | 0,0004 | 0,49 | 167 |
| Masc. | 0,15 | Fem. 0,77 | 0,92 | 3,29 | 0,0011 | 0,0351 | 0,10 | 219 |
| Masc. | 0,97 | Fem. 1,69 | 0,72 | 2,88 | 0,0042 | 0,5130 | 0,23 | 221 |
| Masc. | 1,53 | Fem. 0,86 | 0,67 | 2,61 | 0,0095 | 0,0045 | 0,44 | 56 |
| Masc. | 0,45 | Fem. 1,04 | 0,6 | 2,28 | 0,0231 | 0,1812 | 0,33 | 202 |
| Masc. | 1,78 | Fem. 1,22 | 0,57 | 2,22 | 0,0268 | 0,1779 | 0,44 | 1 |
| Masc. | 0,79 | Fem. 0,2 | 0,59 | 2,17 | 0,0307 | 0,0674 | 0,33 | 91 |
| Masc. | 0,45 | Fem. 1,2 | 0,75 | 2,11 | 0,0356 | 0,0804 | 0,39 | 47 |
| Masc. | 0,34 | Fem. 1,06 | 0,71 | 2,07 | 0,0389 | 0,2268 | 0,81 | 193 |
| Masc. | 0,6 | Fem. 1,13 | 0,53 | 2,06 | 0,0406 | 0,0556 | 0,54 | 166 |

*Graus de liberdade=329

** Mantel-Haenszel

Em decorrência, pode-se considerar que o acordo entre os dois procedimentos foi alto e porcentagem de acordo foi muito alta: um erro de classificação em 211. Esse dado sugere que ambos os procedimentos detectaram a presença

de DIF nos mesmos itens. Infelizmente dada a pouca presença de itens com DIF não foi possível avaliar a acuidade para detectar itens com DIF uniforme e não uniforme.

À GUIA DE CONCLUSÃO

No caso de um instrumento de reconhecimento de palavras, as implicações sociais, educacionais e de tratamento psicológico como consequência do fato de uma classificação que possa informar mais problemas em homens que mulheres, ou vice-versa, são bastante sérias, não apenas em defesa de direitos iguais para ambos os sexos, mas por estereótipos, tratamentos mal encaminhados e decisões educacionais equivocadas. Assim, a criação de instrumentos que controlem esse efeito parece de suma importância, pois o que realmente deverá estar em julgamento é a capacidade e não o sexo da pessoa.

Nesse sentido, os itens selecionados para o início da construção do Teste de Reconhecimento de Palavras mostrou-se uma medida isenta desse problema, possibilitando que se construa um instrumento com base em uma evidência uma validade. Nesse sentido, a pontuação a ser obtida pelo instrumento poderá ser interpretada sem o risco de se estar valorizando característica de um sexo em detrimento do outro. Entretanto, essa mesma análise deverá ser feita depois que o instrumento tiver os itens definidos em razão de outros critérios, pois uma aplicação em larga escala é necessária para uma melhor medida da estabilidade dessa tendência.

Deve-se ressaltar ainda que, embora exista um bom número de procedimentos bastante sofisticados para detectar funcionamento diferencial do item em suas distintas manifestações (Fidalgo, 1996), esses procedimentos são apropriados apenas para detectar o viés em potencial de um item. Eles não possibilitam uma explicação de suas causas (Mellenbergh, 1989), pois sua interpretação deve ser o resultado de uma conjunção entre procedimentos estatísticos e análises teóricas ou julgamento de peritos no assunto. Na realidade, detectar o funcionamento diferencial de um item não possibilita afirmar que ele é enviesado, pois é necessário complementar essa informação relacionando-a com as causas ou razões de seu funcionamento diferenciado em tal grupo e se essas diferenças são ou não parte legítima do constructo em questão. Em decorrência, o exame do conteúdo do item no qual se encontrou evidência de DIF é de suma importância.

Os resultados obtidos ressaltam que, mesmo existindo diferenças reais em relação à

habilidade para reconhecer palavras entre meninos e meninas em fase inicial de escolarização, antes de se comparar e tirar conclusões sobre os dados, é necessário realizar estudos com vistas a uma purificação da escala, para que não seja o viés do item o que determine essa diferença, mas as reais diferenças de habilidade de cada grupo. E essa é uma tarefa, enfatiza-se, deve ser realizada durante a construção do instrumento.

REFERÊNCIAS

- American Educational Research Association [AERA]; American Psychological Association [APA]; and National Council on Measurement in Education [NCME]. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Assink, E.M. H. & Kattenberg, G. P. A (1995). The use of phonological and orthographic information by normal and poor readers of Dutch. *Reading and Writing*, 7(3), 277-294.
- Badian, N. A. (1993). Phonemic awareness, naming, visual symbol processing, and reading. *Reading and Writing*, 5(1), p. 87-100.
- Bartholomeu, D., Sisto, F. F., Rueda, F. J. M.(2006). Dificuldades de aprendizagem na escrita e características emocionais de crianças. *Psicologia em estudo*, 11, 139-146.
- Bazi, G. A. P., Sisto, F. F.(2006). Alegria, tristeza, medo e coragem em crianças com dificuldades de aprendizagem. Em F.F. Sisto & S.C. Martinelli (Orgs.) *Afetividade e dificuldades de aprendizagem*. São Paulo: Vetor Editora, p. 57-72.
- Calhoun, M. L. (1985). Typing contrasted with handwriting in language arts instruction for moderately mentally retarded students. *Education and Training of the Mentally Retarded*, 20(1), 48-52.
- Carvalho, R. M. M., Martes, A. S. H., Camargo, R. L., Chiari, B. M. & Vieira, R. M. (1988). Taxa de redundância na incidência de vogais e consoantes no português escrito. *Acta AWHO*, 7(4), 236-42.
- Cunningham, A. E. & Stanovich, K. E. (1993). Children's literacy environments and early word recognition subskills. *Reading and Writing*, 5 (2), 193-204.

- Draba, R. E. (1977). The Identification and Interpretation of Item Bias. *Rasch Measurement Transactions*, MESA Memorandum no. 25, Disponível em: <http://www.rasch.org/rmt/rmt122m.htm>. Consultado em abril de 2004.
- Fidalgo, A.M. (1996). Funcionamiento diferencial de los ítems. En J. Muñiz (Org.), *Psicometría*. Madrid: Universitas.
- Flanagan, D. P., McGrew, K. S., Abramowitz, E., Untiedt, S. & cols. (1997). Improvement in academic screening instruments? A concurrent validity investigation of the K-FAST, MBA, and WRAT-3. *Journal of Psychoeducational Assessment*, 15(2), 99-112.
- Freyberg, P. S. (1970). The concurrent validity of two types of spelling test. *British Journal of Educational Psychology*, 40(1), 68-71.
- Galdo, N. J. I. (1989). Avaliação precoce: diagnóstico prescritivo. *Arquivos Brasileiros de Psicologia*, 41(4), 71-87.
- Gallagher, A. M., Laxon, V., Armstrong, E. & Frith, U. (1996) Phonological difficulties in high-functioning dyslexics. *Reading and Writing*, 8(6), 499-509.
- Garcia de la Cadena, C., Ostrosky Solis, P., Rodriguez, Y., Madrazo, I. & cols. (1992). Potenciales relacionados a eventos (P300) y procesos cognoscitivos en la enfermedad de Parkinson. *Salud Mental*, 15(4), 50-55.
- Góes, M. C. R. (1984). Noções sobre correspondência som-texto em crianças iniciando a alfabetização. *Arquivos Brasileiros de Psicologia*, 1(36), p. 59-79.
- Gottardo, A., Chiappe, P., Siegel, L. S. & Stanovich, K. E. (1999). Patterns of word and nonword processing in skilled and less-skilled readers. *Reading and Writing*, 11(5-6), 465-487.
- Gough, P. B. (1993). The beginning of decoding. *Reading and Writing*, 5(2), 181-192.
- Graminha, S. S. V., Machado, V. L. S., Francischini, E. L. & Befi, V. M. (1987). Emprego de um procedimento de treino gradual de discriminação de sílabas em crianças com dificuldades na leitura e na escrita. *Arquivos Brasileiros de Psicologia*, 39(1), 84-94.
- Holland, P. W. & Thayer, D. T (1988). Differential item performance and the Mantel-Haenszel procedure. Em H. Wainer & H. I. Braum (Orgs.), *Test Validity*. Hillsdale, NJ: Lawrence Erlbaum.
- Hung, D. L. & Tzeng, O. J. (1981). Orthographic variations and visual information processing. *Psychological Bulletin*, 90(3), 377-414.
- Jaeger, A., Schosler, T. & Wainer, R. (1998). Estudo comparativo da aquisição da escrita em crianças e em adultos. *Psicologia Reflexão e Crítica*, 11(3) 551-558.
- Kajihara, O. (1997) Avaliação das habilidades fonológicas de disléxicos do desenvolvimento. *Tese de doutorado. Instituto de Psicologia da USP*. São Paulo, 272 p.
- Kingenski, M. F., Sisto, F. F.. 2004. Reconhecimento de palavras nas séries iniciais do ensino fundamental. *Teoria e prática da educação*, 7, 173-184.
- Leong, C. K. & Parkinson, M. E. (1992). Sensitivity to orthotactic rules in visual word recognition by below average readers. *Reading and Writing*, 4(1), 1-17.
- Lukatela, K., Carello, C., Shankweiler, D. & Liberman, I. Y. (1995). Phonological awareness in illiterates: Observations from Serbo-Croatian. *Applied Psycholinguistics*, 16(4), 463-487.
- Mahony, D., Singson, M, & Mann, V. (2000). Reading ability and sensitivity to morphological relations. *Reading and Writing*, 12(3-4), 191-218.
- Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719-748.
- Mellenbergh, G.J. (1982). Contingency table models for assesing item bias. *Journal of Educational Statistics*, 7, 105-118.
- Mellenbergh, G.J. (1989). Item bias and item response theory. *International Journal of Educational Research*, 13, 127-143.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35, 1012-1027.
- Messick, S. (1989). Validity. En R.L. Linn (Ed.), *Educational Measurement* (3th. Ed.). New York: American Council on Education and Macmillan publishing company.
- Moss, P. A., Cole, N. S. & Khampalikit, C.A. (1982) comparison of procedures to assess written language skills at Grades 4, 7, and 10. *Journal of Educational Measurement*, 19 (1), 37-47.
- Mota, M., Moussatche, A. H., Castro, C. R. & cols. (2000). Erros de escrita no contexto: uma análise na abordagem do processamento da

- informação. *Psicologia Reflexão e Crítica*, 13(1), 01-06.
- Motomura, N., Tomota, Y., Akagi, H. & Seo, T. (2000). A study of language disorders associated with dementia of Alzheimer type in Japanese: A preliminary study. *Psychologia: An International Journal of Psychology in the Orient*, 43(1), 84-89.
- Oliveira, Q. L. (1996). Três instrumentos de avaliação de habilidades para aprendizagem da leitura e escrita. *Psicologia: teoria e pesquisa*, 12(1), 83-96.
- Pestun, M. S. V. (2001). Análise funcional discriminativa em dislexia do desenvolvimento. *Tese de doutorado. Faculdade de Ciências Médicas da Unicamp. Campinas*, 234 p.
- Pinheiro, A. M. V. & Rothe-Neves, R. (2001). Avaliação cognitiva de leitura e escrita: as tarefas de leitura em voz alta e ditado. *Psicologia Reflexão e Crítica*, 14(2), 399-408.
- Pinheiro, A. M. V. (1995). Dificuldades específicas de leitura: a identificação de déficits cognitivos e a abordagem do processamento de informação. *Psicologia Teoria e Pesquisa*, 11(2), 107-15.
- Rapcsak, S. Z., Rubens, A. B. & Laguna, J. F. (1990). From letters to words: Procedures for word recognition in letter-by-letter reading. *Brain and Language*, 38(4), 504-514.
- Rey, A., Ziegler, J. C. & Jacobs, A. M. (2000). Graphemes are perceptual reading units. *Cognition*, 75(1), 1-12.
- Rueda, F. J. M., Bartholomeu, D., Sisto, F. F. (2004). Desenho da Figura Humana e Aprendizagem da Escrita. Em C. Machado; L. S. Almeida; M. Gonçalves & V. Ramalho (Orgs), *Avaliação Psicológica: formas e contextos*. Braga - Portugal: Psiquilíbrios Edições, p. 236-242.
- Santos, A. A. A. (2002). A influência da consciência fonológica na aquisição da leitura e escrita. Em: F. F. Sisto, G. C. Oliveira, L. D. T. Fini, M. T. C. C. Souza & R. P. Brenelli (Orgs.) *Atuação Psicopedagógica e aprendizagem escolar*. Petrópolis: Editora Vozes, p. 213-247. 7a. edição.
- Shanahan, T. (1984). Nature of the reading-writing relation: An exploratory multivariate analysis. *Journal of Educational Psychology*, 76(3), 466-477.
- Sisto, F. F. (2001). Dificuldade de 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*. Petrópolis: Editora Vozes, p. 190-213.
- Sisto, F. F. (2001). Dificuldades de aprendizagem. Em F.F.Sisto, E.Boruchovitch, L.D.T.Fini, R. P. Brenelli & S.C.Martinelli (Orgs.) *Dificuldades de aprendizagem no contexto psicopedagógico*. Petrópolis: Editora Vozes, p. 19-39.
- Sisto, F. F., Boruchovitch, E., Brenelli, R. P., Fini, L. D. T., Martinelli, S. C., Urquijo, S. (2001). A integridade do ego e o desempenho na escrita. Em F.F.Sisto, E.Boruchovitch, L.D.T.Fini, R. P. Brenelli & S.C.Martinelli (Orgs.) *Dificuldades de aprendizagem no contexto psicopedagógico*. Petrópolis: Editora Vozes, p. 122-147.
- Sisto, F. F., Fernandes, D. C. (2004). Dificuldades lingüísticas na aquisição da escrita e agressividade. *Psicologia Escolar e Educacional*, 8, 75-84.

Recebido em Julho de 2006

Reformulado em Julho de 2006

Aceito em Agosto de 2006

SOBRE O AUTOR:

Fermínio Fernandes Sisto: doutor pela Universidad Complutense de Madrid, Livre – docente pela Unicamp e docente do curso de Psicologia e do Programa de Estudos Pós- graduados em Psicologia, da Universidade São Francisco, campus Itatiba-SP.