

Aplicação da TRI em uma Medida de Avaliação da Compreensão de Leitura

Use of the Item Response Theory on a Measure for Reading Comprehension Assessment

Lucas de Francisco Carvalho^a, Rebecca de Magalhães Monteiro^b,
Adriana Rosecler Alcará^b & Acácia Aparecida Angeli dos Santos^{*, b}

^aUniversidade Presbiteriana Mackenzie, São Paulo, Brasil & ^bUniversidade São Francisco, Itatiba, Brasil

Resumo

Este trabalho objetivou verificar os parâmetros dos itens e dos sujeitos, por meio da Teoria de Resposta ao Item (TRI), em uma medida de avaliação da compreensão de leitura, englobando análises quantitativas e qualitativas do mapa de itens, assim como investigar a presença de funcionamento diferencial dos itens (DIF). Participaram 518 crianças do 3º, 4º e 5º anos do ensino fundamental, com idades entre 6 e 16 anos, de escolas particular e pública de Belo Horizonte. Utilizou-se um texto elaborado de acordo com a técnica de Cloze. Foi confirmada a unidimensionalidade do instrumento; verificou-se média de *theta* maior que a média de dificuldade dos itens; e, a presença de DIF foi observada em alguns itens de acordo com os anos de ensino. Tais resultados demonstraram evidências de validade para o instrumento e são discutidos no trabalho.

Palavras-chave: Compreensão de leitura, Psicometria, ensino fundamental, avaliação psicológica.

Abstract

The objective of the present study was to verify the parameters of items and people by using the Item Response Theory (IRT) in a reading comprehension measurement, including quantitative and qualitative analyses of the items map as well as to investigate the presence of Differential Item Functioning (DIF). The sample consisted of 518 children from the 3rd, 4th and 5th grades, aged from 6 to 16, from private and public schools in the city of Belo Horizonte-MG. The instrument was a text prepared according to Cloze technique. The data confirmed the unidimensionality of the instrument; showed average theta higher than the average of items; and, the presence of DIF was observed in some items in relation to the school grades. The results demonstrated validity evidence for the instrument and are discussed in this paper.

Keywords: Reading comprehension, Psychometric, elementary school, psychological assessment.

Uma das questões relevantes no contexto escolar e que tem preocupado estudiosos de vários países é a defasagem no aprendizado da leitura (Muth, 1989; Olson, 1990; Salles & Parente, 2007). Sabe-se que essa é uma habilidade requisitada em todas as fases da escolarização já que relaciona-se com a aprendizagem de conteúdos diversos (Guidetti & Martinelli, 2007; Oliveira, Boruchovitch, & Santos, 2009; Witter, 2010). Nesse sentido, muitos estudos apontam a importância de se avaliar a leitura no contexto educacional, dada a sua relevância quanto ao desempenho em outras disciplinas curriculares (Joly, 2009; Santos, Primi, Taxa, & Vendramini, 2002; Wang, Jiao, Young, Brooks, & Olson, 2008).

A aquisição da habilidade de leitura envolve vários processos cognitivos, tais como, a decodificação, o desenvol-

vimento da competência linguística, que está, por sua vez, relacionada à apreensão e o desenvolvimento adequado do vocabulário, da sintaxe, da morfologia e da fonologia (Alliende & Condemarin, 1987; Lopes & Paula, 2008). Vale a ressalva de que o ato de ler pode ser compreendido como um processo, no qual a interpretação do que é lido depende, não só do que está impresso, mas também das hipóteses do próprio leitor, formuladas com base no seu conhecimento prévio (Santos, 2004; Santos et al., 2002).

Nessa mesma direção, Kintsch (1994) e Kintsch e Dijk (1978) já consideravam que a compreensão de leitura dependia da interrelação entre vários processos cognitivos. Em seus estudos os autores apontavam que apenas processos básicos como o reconhecimento de palavras e extração do significado das palavras impressas não eram suficientes para a compreensão de um texto. Isso significa dizer que processos cognitivos de alto nível, como a capacidade de realizar inferências, as habilidades de memória e o conhecimento de mundo, facilitam a representação macroestrutural do texto.

* Endereço para correspondência: Universidade São Francisco – Campus Itatiba, Programa de Pós-Graduação Stricto Sensu em Psicologia, Alexandre Rodrigues Barbosa, 45, Itatiba, SP, Brasil 13251-900. E-mail: luca@labape.com.br

Autores como Eysenck e Keane (2007) e Sternberg (2008) apontam que a compreensão de leitura requer vários processos perceptuais e outros cognitivos, assim como um bom conhecimento da língua e da gramática. Dentre os processos perceptuais pode-se citar a percepção das letras, seus formatos e ligações, bem como aspectos relacionados à ortografia. Já os processos cognitivos se caracterizam pela tradução da letra em som, criando um código fonológico, depois a sequência de todos os símbolos visuais e os sons para formar uma palavra relacionando-a a um significado. O conjunto de palavras constituirá uma frase ou sentença, e ao conjunto de sentenças o leitor dará um sentido, passando a entender o texto como um todo. Ao lado disso, os autores ainda consideram que esse é um comportamento cognitivo verbal, sendo que quando o leitor tem contato com um novo conteúdo, o processamento da compreensão se inicia. Também, a memória de trabalho se faz necessária já que permite ao leitor capturar informações anteriores para em seguida associá-las ao novo conteúdo. Nesse sentido, o leitor é capaz de compreender a nova informação armazenando-a no seu banco de memória para utilizá-la em momento posterior.

É importante ressaltar que existem alguns fatores que podem interferir na compreensão de leitura. Dentre os mais frequentes estão o tamanho das sentenças, o uso de palavras incomuns, a complexidade do vocabulário e das ideias contidas no texto e a presença de categorias gramaticais mais difíceis, além do conhecimento prévio e o interesse do leitor pelo assunto (Cabral & Kastrup, 2009; Phakiti, 2003; Santos, 2004; Spinillo & Mahon, 2007; Witter, 2010). Dessa forma, a compreensão de leitura configura-se como um processo interativo que envolve variáveis inerentes ao texto, tais como sua estrutura e a complexidade do vocabulário, como também características do leitor, dentre as quais se destacam o interesse pelo assunto e o conhecimento prévio.

Sobre essas questões, Singer e Donlan (1982) já ressaltavam que a compreensão de leitura consiste na interação entre os recursos do leitor e as características do texto. E para que essa interação aconteça é fundamental que o leitor use o conhecimento prévio de informações semelhantes às abordadas no texto. Assim, a compreensão de leitura também está relacionada aos conceitos, valores e conhecimentos anteriormente assimilados.

Ao refletir sobre a compreensão de leitura, é importante lembrar que ela é um dos fatores determinantes da aprendizagem e do desenvolvimento de um aluno, já que esse recurso de linguagem é utilizado por todas as disciplinas na escola. Nesse sentido, é possível inferir que quando o aluno apresenta um bom repertório de conhecimentos armazenados possivelmente apresentará melhor compreensão, visto que a compreensão de leitura depende de retroalimentação (Gilbert, Martínez, & Vidal-Abarca, 2005).

A avaliação da compreensão de leitura é um processo complexo, pois a sua observação não ocorre de forma direta e objetiva. Assim, avaliar essa habilidade depende

de um instrumento que permita não apenas diagnosticar a capacidade de produção textual do sujeito, mas também sua compreensão real de leitura, incluindo-o como parte dinâmica nesse processo bem como outros aspectos envolvidos nessa compreensão, como sua capacidade de realizar inferências e analogias (Oliveira et al., 2009; Rupp, Ferne, & Choi, 2006). Nessa direção, a técnica de Cloze tem se destacado pela sua utilidade tanto para o diagnóstico como para a intervenção no processo de aprendizagem da compreensão de leitura. Vale destacar que a técnica de Cloze pode ser empregada como ferramenta de avaliação, diagnóstico e intervenção da compreensão de leitura, quando é mais comumente chamada de teste de Cloze.

Esse instrumento na sua forma inicial, tal como proposto por Taylor em 1953, consistia em um texto de 250 palavras, do qual a cada cinco palavras uma era omitida e substituída por um traço proporcional ao seu tamanho. Os resultados de pesquisas com textos que utilizaram a técnica de Cloze indicaram sua capacidade como ferramenta para avaliar os processos cognitivos envolvidos na compreensão de leitura (Guidetti & Martinelli, 2007; Joly, 2009; Santos et al., 2002; Santos & Oliveira, 2010; Taylor, 1953, entre outros).

No que se refere aos itens do teste de Cloze, o nível de dificuldade pode estar relacionado a fatores intrínsecos e extrínsecos. Os primeiros dizem respeito aos elementos positivos que facilitam a execução da tarefa e funcionam como um estímulo para o entendimento do item. Já os extrínsecos incluem os conhecimentos que o estudante tem sobre o assunto e as estratégias utilizadas para o desempenho da tarefa (Abraham & Chapelle, 1992; Oliveira et al., 2009). De acordo com as autoras, ao apreender um texto de Cloze, o leitor age como um processador de textos escritos que, por meio de conhecimentos prévios, faz inferências e analogias que resultam na sua compreensão.

Para melhor ilustrar a avaliação da leitura por meio da técnica de Cloze destacam-se duas pesquisas mais atuais que buscaram investigar as dificuldades dos itens em um texto específico “*A Princesa e o Fantasma*” elaborado por Santos em 2005. Trata-se de uma história infantil, especialmente desenvolvida para crianças do ensino fundamental, contendo 103 palavras estruturadas na forma do Cloze tradicional, proposta por Taylor (1953), com 15 omissões substituídas por um traço proporcional ao seu tamanho a cada 5º vocábulo. Este estudo preliminar foi realizado com 314 alunos de ambos os sexos, de 2ª a 4ª séries (atuais 3º a 5º anos) e por meio da análise da variância identificou diferença estatisticamente significativa entre as pontuações dos alunos das séries, que era mais elevada conforme seu avanço [$F(3,314)=55,75$; $p<0,001$]. O teste *post-hoc* de Tukey, realizado a seguir, separou as pontuações em três agrupamentos, conferindo evidência de validade de critério. A análise da consistência interna apresentou índices de fidedignidade satisfatórios ($>0,70$ $<0,86$; Urbina, 2007).

O estudo de Cunha e Santos (2009) teve como objetivos explorar as diferenças qualitativas nos erros apresentados no mesmo teste de Cloze “*A Princesa e o Fantasma*” com o objetivo de encontrar evidências de validade por processo de resposta. Participaram da pesquisa 266 alunos de 3ª e 4ª série do Ensino Fundamental (atuais 4º e 5º anos), de escolas particulares e públicas de São Paulo, com idades entre 8 e 13 anos. Após as análises descritivas, as autoras separaram protocolos de participantes que constituíram dois grupos extremos, alunos com escores baixos ($n=22$) e alunos com escores altos ($n=22$). Para a avaliação do processo de resposta foi proposta a seguinte classificação dos erros: 1=branco, 2= erro fonológico (incluindo erros ortográficos e de acentuação), 3=erro lexical (uso de sinônimos), 4=erro sintático, 5=erro semântico. Os alunos com escores altos (>12 pontos) cometeram mais erros lexicais, seguidos dos sintáticos e dos semânticos. Na classificação lexical, as palavras que proporcionaram mais erros foram “misterioso” e “do”. Já com os alunos que obtiveram escores baixos (<7 pontos), houve coincidência com a ordem crescente dos totais de erros cometidos, iniciando pelos erros fonológicos, seguidos dos lexicais, dos sintáticos e por fim dos semânticos. Em síntese, os resultados encontrados evidenciaram que crianças com escores altos cometeram mais erros lexicais, enquanto que as com escores baixos mais erros semânticos. Isso significa que erros por substituição de sinônimos podem ser considerados menos prejudiciais do que os erros semânticos.

Ainda usando o mesmo texto de Santos (2005) mencionado anteriormente, Santos, Carvalho, Monteiro e Alcará (2010) se propuseram a verificar os parâmetros dos itens e das pessoas, por meio da Teoria de Resposta ao Item (TRI), englobando também análises quantitativas e qualitativas do mapa de itens. Participaram da pesquisa 170 crianças do 4º ano do ensino fundamental de uma escola pública e uma particular de Minas Gerais, com idade variando entre 8 e 14 ($M=9$ anos; $DP=0,72$). Os resultados apontaram adequação dos parâmetros do instrumento, bem como adequação da amostra para o instrumento e texto utilizados. Em relação às pessoas, uma distinção de dois níveis indicou que os participantes da pesquisa, do ponto de vista estatístico, poderiam ser subdivididos em pelo menos duas categorias distintas (por exemplo, baixo e alto em relação à compreensão de leitura). A respeito dos itens, os *strata* encontrados indicaram que os itens poderiam ser agrupados, estatisticamente, em pelo menos 10 grupos distintos. O item três teve que ser excluído das análises, pois nenhum aluno o respondeu corretamente. O item nove foi considerado o mais fácil para amostra estudada, já que os respondentes com theta igual ou inferior a aproximadamente -2,5 foram os que acertam somente esse item. O item mais difícil foi o 10 ($b=3,94$).

Ao considerar tais estudos é possível destacar a importância de pesquisas que analisem os itens do texto utilizado para compreensão de leitura no nível do próprio

item, isto é, não tratando somente como um conjunto expresso em um escore total. Contudo, não foram encontrados estudos, além de Cunha e Santos (2009) e Santos et al. (2010) que tivessem essa proposta. Sendo assim, o objetivo deste estudo foi verificar os parâmetros dos itens e das pessoas, por meio da Teoria de Resposta ao Item (TRI), de um texto de Cloze para avaliação da compreensão de leitura, analisando quantitativa e qualitativamente o mapa de itens. Além disso, buscou-se verificar a presença de funcionamento diferencial (DIF) nos itens do instrumento por meio da variável ano escolar. Vale ressaltar que este estudo se diferencia do trabalho apresentado por Cunha e Santos (2009) na medida em que utiliza um dos modelos matemáticos com base na TRI, o modelo de Rasch; e do trabalho apresentado por Santos et al. (2010), já que no presente estudo a amostra é composta por crianças do 3º, 4º e 5º ano dos ensino fundamental e foi investigada a possibilidade de funcionamento diferencial dos itens (análise DIF), que será descrita em parágrafo posterior.

Método

Participantes

Foram participantes deste estudo 518 crianças do 3º, 4º e 5º anos do ensino fundamental, com idades variando entre 6 e 16 ($M=9,10$ anos; $DP=1,20$), sendo 54,2% ($N=281$) meninos. Além disso, 363 eram de escola particular e 155 de escola pública. Quanto à distribuição dos alunos por ano escolar 173 cursavam o 3º ano, 170 o 4º ano e 175 o 5º ano. As médias de idade dos alunos de cada um desses anos escolares foram respectivamente ($M=8,16$; $DP=1,00$), ($M=9,00$; $DP=0,72$) e ($M=10,13$; $DP=0,93$).

Instrumentos

Nesta pesquisa utilizou-se a técnica de Cloze a partir do texto “*A Princesa e o Fantasma*”. A maior ou menor facilidade do leitor para reconstruir a estrutura do texto determinará não só o índice de sua inteligibilidade, como também sua habilidade de compreensão (Oliveira et al., 2009).

O texto utilizado no presente trabalho foi a “*A Princesa e o Fantasma*”, elaborado por Santos (2005) tal como já mencionado. Assim, solicitou-se aos participantes a leitura da história que possui alguns espaços em branco para serem completados com palavras que tenham o melhor sentido para o texto. Adotou-se a correção literal ou verbatim, só se aceitando como correta a resposta exatamente igual à que foi omitida, sendo atribuído um ponto por acerto. Dessa forma, a pontuação mínima e máxima foi respectivamente zero e 15 pontos.

Foram encontrados na literatura estudos que utilizaram o texto “*A Princesa e o Fantasma*” a partir da técnica de Cloze (Cunha & Santos, 2008, 2009; Guidetti & Martinelli, 2007; Oliveira et al., 2009; Santos & Oliveira, 2010, entre outros). No geral, esses estudos encontra-

ram evidências de validade como medida de compreensão de leitura, avaliando as relações entre seus resultados e outras habilidades, tais como o desempenho em escrita. Também buscaram diferenças entre os escores globais do Cloze e variáveis critério, como ano escolar, visto que ele aumenta conforme a progressão do ano escolar, o que foi corroborado pelos dados observados. Além disso, de acordo com os dados apresentados por Santos et al. (2010), o teste apresenta consistência interna de 0,83, o que pode ser considerado como um índice adequado de fidedignidade (Urbina, 2007).

Procedimentos e Análise dos Dados

Primeiramente foi solicitada a autorização da direção das escolas para a realização da pesquisa e, em seguida, o projeto foi encaminhado para a aprovação do Comitê de Ética. Os participantes foram informados sobre os objetivos do estudo e sobre a exigência da assinatura do Termo de Consentimento Livre e Esclarecido (TCLE) pelo responsável, já que os alunos eram menores de idade.

As informações sobre a pesquisa foram transmitidas para as crianças e os pais participantes, primeiramente, pelas diretoras das escolas e, posteriormente, no dia da aplicação dos testes pela pesquisadora. Para os pais das crianças foi feito um comunicado por escrito, por meio de uma circular das escolas e junto foi encaminhado o termo de consentimento livre e esclarecido para os pais autorizarem ou não a pesquisa. O texto “*A Princesa e o Fantasma*” foi aplicado coletivamente nos alunos em sala de aula durante o período de aula no horário mais apropriado e liberado pela direção das escolas. O tempo gasto para a realização da tarefa foi de aproximadamente 20 minutos.

Os resultados obtidos foram submetidos às análises do modelo de Rasch, por meio do *software* estatístico Winsteps® versão 3.69.1 (Linacre & Wright, 2009), de modo que foram verificados parâmetros tanto do instrumento quanto dos respondentes. Basicamente, foram realizadas as seguintes análises: estatísticas descritivas do nível de habilidade (*theta*) dos participantes; estatísticas descritivas dos itens; índices de ajuste (*infit* e *outfit*) ao modelo; índices de fidedignidade e de separação; valores dos limiares dos itens; mapa de pessoas-itens; análise quantitativa e qualitativa da divisão dos itens e das pessoas (mapa de itens); e análise do funcionamento diferencial dos itens (DIF) de acordo com o ano escolar dos participantes.

Resultados e Discussão

Para a análise do conjunto de itens foram considerados todos os 15 itens derivados do texto “*A Princesa e o Fantasma*”, de modo que, em um primeiro momento foi verificada a possibilidade de unidimensionalidade desse conjunto de itens. Para tanto, utilizou-se da Análise por Componentes Principais de Resíduos baseada no modelo

de Rasch. Os dados obtidos apontaram para 55,8% de variância explicada pelos itens, equivalente a 18,9 *eigenvalues*, e não foram encontrados *eigenvalues* de tamanho suficiente (isto é, 2,0) para consideração dos contrastes (Linacre & Wright, 2004, 2009). Nesse sentido, pode-se dizer que o conjunto de itens é unidimensional.

Na Tabela 1 podem ser verificadas as estatísticas descritivas sumarizadas acerca dos *thetas* (nível de habilidade) dos respondentes e seus respectivos índices de ajuste (*infit* e *outfit*). A média de *theta* encontrada sugere que, no geral, o conjunto de itens tende a ser marginalmente fácil para a amostra deste estudo, isto é, as crianças tenderam a acertá-los mais do que errá-los. Pode-se dizer que esses dados vão ao encontro aos achados de Cunha e Santos (2009), já que no estudo das autoras, 10 (de 15) itens não apresentaram mais que 4 erros ao todo, apesar de alguns itens terem obtido mais de 10 erros. Também corrobora os dados apresentados por Santos et al. (2010), que evidenciou uma média de *theta* positiva (1,28). A partir disso pode-se pensar no uso de textos para a técnica de Cloze, para a faixa etária e escolar da amostra estudada, contemplando itens com níveis de dificuldades mais altos. Esse dado fica evidente também na Figura 2, apresentada neste estudo, na qual é possível verificar uma lacuna de itens em níveis mais altos de dificuldade.

Além disso, os índices médios de ajuste *infit* e *outfit*, que avaliam a correspondência entre os valores esperados e observados das estimativas *thetas* para os respondentes, mostraram-se adequados, conforme os parâmetros sugeridos por Linacre e Wright (2004), qual seja, inferiores a 1,20. Contudo, os valores máximos de *infit* e *outfit* excederam esse limite, indicando que a pontuação de alguns respondentes não se ajustou adequadamente ao que era esperado pelo modelo. Considerando os objetivos deste estudo, investigar os parâmetros dos itens e dos sujeitos por meio da TRI, os sujeitos (e o mesmo vale para os itens) que apresentaram *infit* e *outfit* elevados não foram retirados da amostra. Ainda assim, a média nesses índices sugere que a maior parte da amostra não apresentou desajuste. Contudo, sugere-se que futuros estudos investiguem possíveis diferenças nos resultados a partir da exclusão de indivíduos que apresentem desajuste nesses índices, bem como itens que apresentem níveis elevados de desajuste.

Em complemento, verificou-se que a fidedignidade das estimativas de *theta* dos respondentes calculada pelo modelo de Rasch foi igual a 0,73 (índice real) e igual a 0,77 (índice dos escores modelados). Pode-se considerar esses índices como satisfatórios, indicando uma quantidade de erro aceitável na avaliação da compreensão de leitura. Também os índices de separação mostraram-se próximo ao que é considerado como satisfatório (Wright & Masters, 1982), sendo 1,45 (índice real) e 1,49 (índice dos escores modelados). O índice de separação refere-se ao número de níveis de distinções estatísticos de severidade entre a amostra incluída na análise. Especificamente,

Tabela 1

Estatísticas Descritivas dos Valores de Theta dos Respondentes e Índices de Ajuste dos Dados

	Número de itens	Theta	Erro Padrão	<i>Infit</i>	<i>Outfit</i>
Média	15	0,60	0,81	0,98	0,99
Desvio Padrão	0	1,70	0,13	0,45	1,50
Máximo	15	5,04	1,45	3,64	9,90
Mínimo	15	-4,02	0,68	0,23	0,05

Tabela 2

Estatísticas Descritivas dos Índices de Dificuldade e Ajuste dos Itens

	<i>N</i>	<i>b</i>	Erro Padrão	<i>Infit</i>	<i>Outfit</i>
Média	505,3	0	0,15	1,00	1,60
Desvio Padrão	0,7	2,42	0,06	0,09	2,27
Máximo	507	5,93	0,37	1,17	9,90
Mínimo	505	-2,73	0,11	0,80	0,49

esse índice descreve a variância “verdadeira” em unidade de variância de “erro” (Myford & Wolfe, 2004). A partir dos índices de separação são calculados o *strata*, que é usado para determinar o número de níveis de distinções estatísticas da habilidade das pessoas que os itens distinguiram (Smith, 2001). Na Tabela 2 encontram-se as estatísticas descritivas sumarizadas dos itens do instrumento.

O índice de dificuldade dos itens variou de -2,73 a 5,93, com desvio padrão igual a 2,42. Esses dados se referem a amplitude de cobertura dos itens ao construto em relação às pessoas, e poderão ser melhor observados posterior-

mente no mapa de itens com maior detalhe (Figura 1). A média do índice de ajuste dos itens, *infit*, mostrou-se adequada, isto é, abaixo de 1,20, mas a média de *outfit* apresentou índice superior ao desejado. Além disso, os índices máximos de *infit* e *outfit* sugerem que pelo menos um item apresentou incongruência entre o esperado pelo modelo e o observado nos dados. Os índices de fidedignidade foram iguais a 1,00 (real e modelado), e os índices de separação foram iguais a 14,14 (índice real) e 14,64 (índices de escores modelados). Na Tabela 3 são encontrados os valores médios de dificuldade para cada item.

Tabela 3

Índices de Dificuldade, Ajuste e Correlação Item-Total

Itens	Número de Erros (Cunha & Santos, 2009)	<i>B</i>	Erro Padrão	<i>Infit</i>	<i>Outfit</i>	Correlação Item-Total
3	17	5,93	0,37	1,17	9,90	-0,04
10	14	3,18	0,15	0,98	1,13	0,40
2	11	2,34	0,12	1,09	1,11	0,43
14	7	2,00	0,12	1,00	2,61	0,46
7	12	1,41	0,11	1,05	0,97	0,52
4	4	0,22	0,11	0,99	0,91	0,58
11	3	-0,49	0,12	0,84	0,70	0,65
5	0	-0,51	0,12	0,97	0,79	0,60
13	1	-1,08	0,13	1,08	0,99	0,53
15	3	-1,34	0,13	0,99	0,98	0,55
6	4	-1,86	0,15	1,08	1,24	0,48
8	1	-2,08	0,15	0,97	0,78	0,52
12	0	-2,35	0,16	0,80	0,49	0,57
1	2	-2,64	0,17	0,96	0,70	0,50
9	1	-2,73	0,18	1,00	0,73	0,47

De acordo com a Tabela 3, o item mais fácil foi o 9 ($b = -2,73$) e o mais difícil foi o 3 ($b = 5,93$), sendo que a maior parte dos itens obteve índice de dificuldade negativo. Índices de dificuldade negativos sugerem que os itens apresentam, nesta amostra, uma tendência para serem acertados mais que respondidos de maneira incorreta. Comparando o nível de dificuldade dos itens com os números de erros apresentados por Cunha e Santos (2009), observa-se uma concordância, isto é, em geral itens mais difíceis neste estudo foram aqueles que mais apresentaram erros no estudo das autoras e, da mesma maneira, os itens presentemente mais fáceis, foram os que obtiveram menos erros no estudo citado. Os índices de ajuste apresentaram-se bastante similares, de modo que somente alguns itens (*oufit*: 3, 6 e 14) obtiveram índices mais altos do que é esperado, especialmente o item 3. Considerando que no estudo de Santos et al. (2010) nenhum aluno respondeu corretamente o item 3, isto é, não houve variância nas respostas dadas ao item, o mesmo foi excluído das análises.

No caso do presente estudo o item 3 obteve correlação item-total negativa, sugerindo que, além de não contribuir com a fidedignidade do conjunto de itens, prejudica a mesma. O índice de *oufit* alto e a correlação negativa são pontos que sugerem a inadequação desse item na avaliação da compreensão em leitura. É possível dizer que esse resultado corrobora o achado de Santos et al. (2010). Com exceção do item 3, a correlação item-total variou entre 0,40 e 0,65, apontando para uma tendência aos itens em avaliar o mesmo construto. Na continuidade é apresentado o mapa de pessoas-itens, Figura 1.

Verifica-se que os itens do instrumento permeiam grande parte do construto compreensão de leitura em relação ao nível de habilidade dos respondentes. Desse modo, apenas pequenas lacunas na mensuração do construto podem ser identificadas, sobretudo próximo a média e em níveis extremos, altos e baixos. Complementando os dados apresentados até este ponto, e auxiliando na compreensão dos escores no conjunto de itens presentemente avaliados, observa-se na Figura 2 o mapa de itens. Ainda é possível observar, pela Figura 1, a discrepância do item 3 em relação aos demais itens do instrumento, o que justifica o fato de talvez excluí-lo das análises neste estudo em especial assim como Santos et al. (2010). Além disso, pode-se inferir que por ser um adjetivo, *misterioso*, demanda um vocabulário mais amplo e talvez essa palavra não fosse muito familiar para as crianças deste estudo.

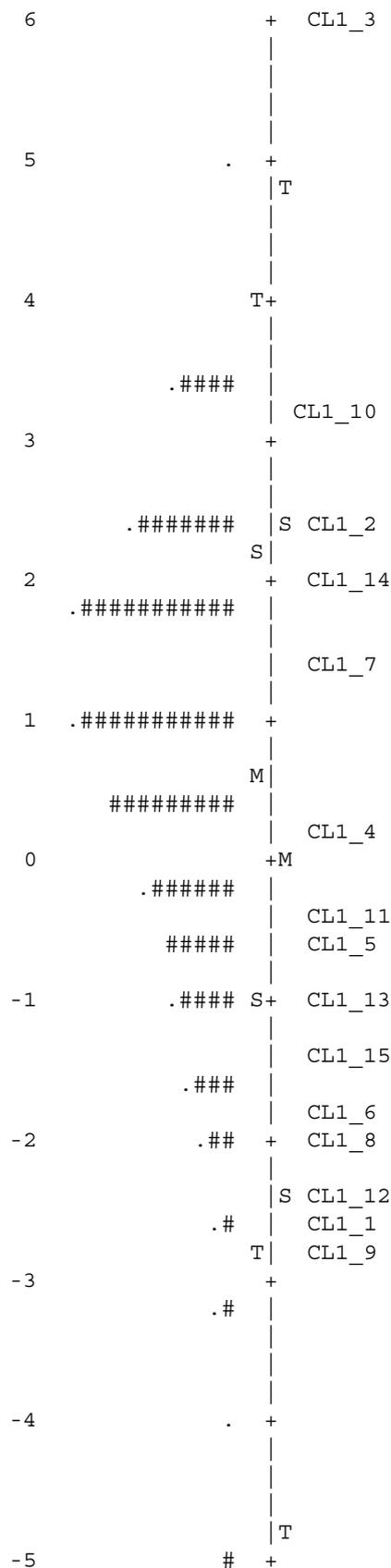


Figura 1. Relação do nível de habilidade e dificuldade por meio do Mapa de pessoas-itens.
 Nota. Legenda: “#” equivalente a 7 pessoas e “.” equivalente a até 6 pessoas.

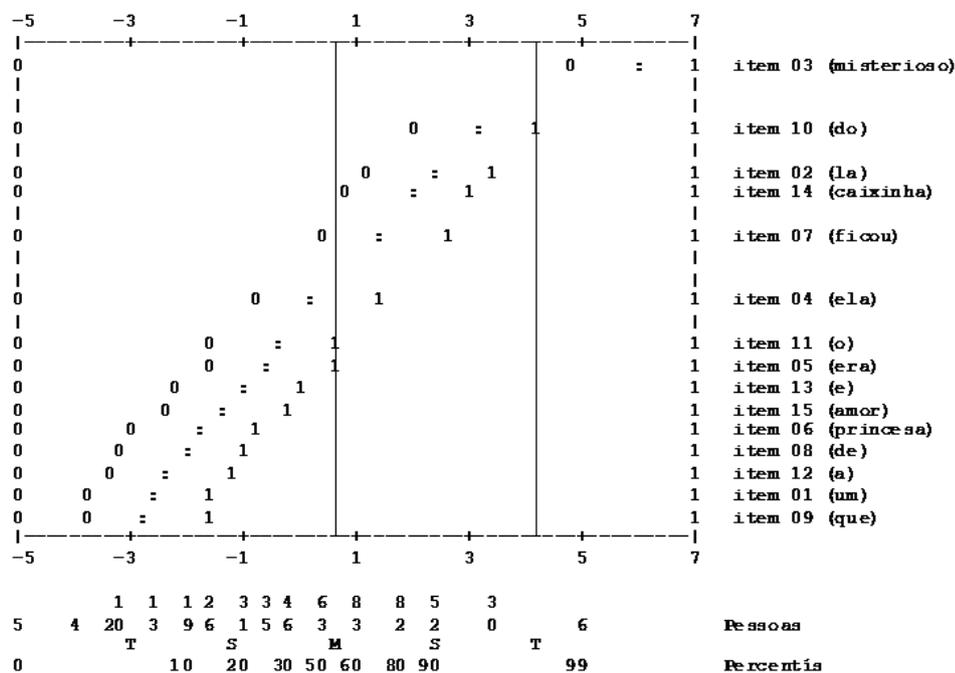


Figura 2. Categorias de respostas no Mapa de itens.

Os itens são apresentados, no mapa, dos mais fáceis (base) para os participantes até os mais difíceis (topo). Para cada item são apresentados o número e a descrição. No corpo do mapa são exibidas as categorias de resposta (0 para erro e 1 para acerto) para cada item, cuja posição é dada pelos limiares na escala de *theta*. Na margem inferior do mapa é apresentada a distribuição dos respondentes (o número de respondentes em cada faixa deve ser lido na vertical), a escala de *theta* (que variou de -5 a +7), e os percentis da amostra. Logo abaixo da frequência de participantes estão dispostas as letras T, S e M, que se referem, respectivamente: dois desvios padrões (abaixo ou acima da média), um desvio padrão (abaixo ou acima da média), e a média.

De acordo com os índices de separação encontrados para pessoas e itens, apresentados anteriormente, após convertidos em *strata*, isto é, índices que indicam o nível de distinções estatísticas (Elliott et al., 2006; Smith, 2001; Wright & Masters, 1982), as pessoas podem ser divididas em 2 (índice real e modelado) níveis e os itens em 20 (índice real e modelado) níveis. Em relação às pessoas, uma distinção de 2 níveis indica que os participantes da pesquisa, do ponto de vista estatístico, podem ser subdivididos em pelo menos duas categorias distintas (por exemplo, baixo e alto em relação à compreensão de leitura). A respeito dos itens, os *strata* encontrados indicam que os itens exibem distinções estatísticas equivalentes a 20 grupos distintos. É desejável obter o maior número possível de distinções estatísticas entre os itens, pois esse dado funciona como indicativo de que os diversos itens, em diferentes níveis de dificuldade, são capazes de discriminar sujeitos com diferentes níveis de habilidade

(*thetas*). Para o presente estudo, considerando os *strata* e uma análise qualitativa da Figura 2, foram estabelecidos dois pontos, derivando três distinções de níveis de habilidade, na tentativa de exemplificar o uso prático do instrumento.

A primeira reta vertical da esquerda (estabelecida a partir da média dos sujeitos) separa respondentes com *theta* igual ou inferior a 0,60, que acertam o conjunto de itens que varia entre o item 9 (o mais fácil) e o item 11. Já a segunda reta, localizada no segundo desvio padrão acima da média, aponta para respondentes que acertam além dos itens anteriores, pelo menos mais cinco itens (4, 7, 14, 2, e 10). E, o último grupo separado pelas duas retas, corresponde aos indivíduos que tendem a acertar, além dos já citados, também o item de número 3 (o mais difícil). Algumas considerações desses agrupamentos podem ser realizadas.

Por exemplo, nota-se que as distâncias entre cada item do grupo que varia entre o 4 e o 10 é mais distinta que as distâncias verificadas entre cada item do grupo que vai do 9 ao 11. Esse dado sugere que o grupo de itens variando entre o 4 e o 10 tende a ser mais discriminativo para a amostra do que o outro grupo citado. Além disso, apesar do item 3 exibir uma distância bastante clara do restante dos itens, vale a pena notar que ele é um item muito difícil, de modo a ser útil para casos de pessoas com níveis de habilidades muito altos no construto avaliado. Nesse sentido, o número escalar padronizado em *theta* não representa apenas um número arbitrário na escala, mas é possível inferir quais habilidades o sujeito com um dado nível de habilidade possui e quais não (Embretson, 2006). Entretanto, há uma lacuna importante entre os itens 10 e

03, isto é, não há itens avaliando a compreensão de leitura naquele nível do construto, a partir da técnica e texto utilizados nesta amostra. Esse dado sugere a necessidade de inclusão de itens avaliando níveis mais extremos do construto compreensão de leitura para crianças com características similares às presentemente avaliadas.

O mesmo ocorreu no estudo de Santos et al. (2010), evidenciando assim a aplicabilidade deste instrumento no contexto educacional. Sob essa perspectiva, vale lembrar que a leitura é uma habilidade requisitada durante todo o período escolar por diversas disciplinas, portanto sua avaliação se torna necessária a fim de identificar alunos tenham dificuldade em adquirir tal habilidade (Guidetti & Martinelli, 2007; Joly, 2009; Salles & Parente, 2007; Santos et al., 2010; Spinillo & Mahon, 2007; Wang et al., 2008).

Por último, são apresentados os resultados relativos ao funcionamento diferencial do item (DIF), de acordo com a série dos participantes da pesquisa. Para tanto, foram utilizados dois critérios distintos para considerar a existência ou não de DIF nos itens, já que não há consenso sobre esse ponto (Du, 1995; Primi, Carvalho, Miguel e Silva, 2010). Um dos critérios foi o de Draba (1977), que considera a presença de DIF em itens cuja diferença entre as dificuldades dos itens refletem um escore t igual ou superior a 2,4. O outro critério é o proposto em alguns trabalhos (Linacre & Wright, 2009; Tennant & Pellant, 2007), que estabelece a presença de DIF de acordo com o *logit*, isto é, terá DIF quando a diferença entre as dificuldades de um mesmo item para dois grupos distintos for igual ou superior a 1,5. A Tabela 4 apresenta os dados relativos ao DIF.

A tabela é composta por 5 colunas, sendo a primeira relativa aos quinze itens (considerando três linhas para cada item). A segunda coluna refere-se a que grupos estão sendo comparados, por exemplo, 3-4 significa que os sujeitos da terceiro ano e quarto ano estão sendo comparados naquela linha. A terceira e quarta colunas informam acerca da presença ou não de DIF de acordo com os critérios descritos anteriormente. E, a última coluna diz respeito à que grupo foi privilegiado quando da presença do DIF. As células preenchidas com “—” indicam a ausência de DIF para o item na comparação de dois grupos determinados.

De acordo com os dados apresentados, observa-se que os itens 3 e 7 foram aqueles que obtiveram DIF de acordo com os dois critérios estabelecidos em duas comparações (das 3 possíveis), e o item 9 apresentou uma categoria que preencheu ambos os critérios. Ao lado disso, alguns itens (1, 4, 11, 12, 15) apresentaram DIF somente de acordo com o índice apresentado em *logits* (Linacre & Wright, 2009; Tennant & Pallant, 2007), e nenhum dos itens apresentou DIF somente no critério proposto por Draba (1977), sugerindo que esse é um critério mais rigoroso em relação ao de *logits*. E, alguns itens não apresentaram DIF em nenhuma das comparações, sendo eles, 2, 5, 6, 8, 10, 13, e 14.

Em relação aos grupos privilegiados na presença do DIF, foram observadas algumas inconsistências, isto é, casos em que o grupo privilegiado é o grupo de pessoas supostamente menos habilidosas (no caso, a relação entre série e habilidade seria positiva). Por exemplo, no item 1, quando a comparação dos grupos 4 (quarto ano) e 5 (quinto ano), o grupo 4 é favorecido pelo DIF. A presença de DIF é indicativo de que um segundo fator, que não o principal construto avaliado (compreensão de leitura), esteja influenciando a resposta do indivíduo àquele item (Linacre & Wright, 2009).

Seria necessária uma investigação mais aprofundada nesse sentido, na tentativa de verificar qual seria esse fator, isto é, se são atributos mais relacionados ao sujeito (por exemplo, conhecimento de vocábulos específicos, memória semântica, entre outros), ao item (por exemplo, localização do item no texto, entre outros), ou ambos. Este achado psicométrico é congruente com o mencionado por Abraham e Chapelle (1992) sobre a interferência de fatores específicos na compreensão de textos. A esse respeito os autores alertam que a predição da dificuldade do item pode depender do local da pista do contexto, do número de sílabas no período, da extensão da palavra a ser recuperada, da quantidade de respostas possíveis para a lacuna, bem como do número de alternativas a serem consideradas. Além disso, seria importante verificar o impacto prático do DIF, ou seja, investigar a necessidade (ou não) de desenvolver tabelas distintas para os itens que apresentaram DIF de acordo com os diferentes anos escolares.

Considerações Finais

Este estudo teve como objetivo verificar os parâmetros dos itens e das pessoas, por meio da Teoria de Resposta ao Item (TRI), de um texto de Cloze para avaliação da compreensão de leitura, o que incluiu análise da presença de DIF nos itens. No geral, os dados apontaram adequação dos parâmetros do instrumento, bem como adequação da amostra para o instrumento e texto utilizados. Ainda assim, alguns sujeitos e itens apresentaram índices elevados de *infit* e *outfit* e, especificamente em relação aos itens, foi evidenciado o fenômeno do DIF. Como já apontado anteriormente, futuros estudos devem aprofundar na investigação dessas questões.

Ao lado disso, os dados encontrados corroboram os achados de Cunha e Santos (2009) e Santos et al. (2010), sugerindo consistência dos resultados evidenciados. É importante notar que a consistência do funcionamento do instrumento foi encontrada apesar das diferenças entre as análises utilizadas (por exemplo, TCT e TRI) e as diferenças na amostra (por exemplo, crianças do 4º ano e crianças do 3º ao 5º ano).

Vale apontar que não foram encontrados estudos em âmbito nacional que utilizassem a Teoria de Resposta ao Item (TRI) para avaliar a compreensão de leitura. Bus-

Tabela 4

Presença de DIF nos Itens de acordo com os Critérios

Itens	Grupos Comparativos	DIF (Draba, 1977)	DIF (Linacre & Wright, 2009; Tennant & Pallant, 2007)	Grupo Privilegiado
1	3-4	—	—	—
1	3-5	—	—	—
1	4-5	Não	Sim	4
2	3-4	—	—	—
2	3-5	—	—	—
2	4-5	—	—	—
3	3-4	Sim	Sim	3
3	3-5	Sim	Sim	3
3	4-5	Não	Sim	5
4	3-4	—	—	—
4	3-5	Não	Sim	5
4	4-5	—	—	—
5	3-4	—	—	—
5	3-5	—	—	—
5	4-5	—	—	—
6	3-4	—	—	—
6	3-5	—	—	—
6	4-5	—	—	—
7	3-4	Sim	Sim	3
7	3-5	Sim	Sim	3
7	4-5	—	—	—
8	3-4	—	—	—
8	3-5	—	—	—
8	4-5	—	—	—
9	3-4	Sim	Sim	4
9	3-5	—	—	—
9	4-5	Não	Sim	4
10	3-4	—	—	—
10	3-5	—	—	—
10	4-5	—	—	—
11	3-4	—	—	—
11	3-5	Não	Sim	5
11	4-5	Não	Sim	5
12	3-4	Não	Sim	4
12	3-5	Não	Sim	5
12	4-5	—	—	—
13	3-4	—	—	—
13	3-5	—	—	—
13	4-5	—	—	—
14	3-4	—	—	—
14	3-5	—	—	—
14	4-5	—	—	—
15	3-4	—	—	—
15	3-5	—	—	—
15	4-5	Não	Sim	5

cou-se, então, mais do que apresentar a adequação do teste de Cloze “*A Princesa e o Fantasma*” sob a perspectiva da TRI, ilustrar o potencial da TRI para análises de com-

preensão de leitura e construtos relacionados ao desempenho acadêmico. Dessa forma, pode ser uma ferramenta importante para o estudo de variáveis do contexto edu-

cacional, contribuindo para a identificação fácil, rápida e de baixo custo para identificar dificuldades de compreensão da leitura, favorecendo o uso de práticas educativas preventivas na própria sala de aula, com textos escritos que estejam acessíveis.

Referências

- Abraham, R. G., & Chapelle, C. A. (1992). The meaning of Cloze test scores: An item difficulty perspective. *The Modern Language Journal*, 76(4), 468-479.
- Alliende, F., & Condemarín, M. (1987). *Leitura – Teoria, avaliação e desenvolvimento* (J. C. A. Abreu, Trad.). Porto Alegre, RS: Artes Médicas.
- Cabral, M. C. C., & Kastrop, V. (2009). Leitura de acolhimento: Uma experiência de devir consciente. *Psicologia: Reflexão e Crítica*, 22(2), 286-293.
- Cunha, N. B., & Santos, A. A. A. (2008). Habilidades lingüísticas no ensino fundamental em escolas públicas e particulares. *PSIC - Revista de Psicologia da Vetor Editora*, 9(1), 35-44.
- Cunha, N. B., & Santos, A. A. A. (2009). Validade por processo de resposta no teste de Cloze. *Fractal: Revista de Psicologia*, 21(3), 549-562.
- Draba, R. E. (1977). *The Identification and Interpretation of Item Bias* (Research Memorandum No. 25). Chicago, IL: Statistical Laboratory, Department of Education, University of Chicago. Retrieved April 17, 2004, from <http://www.rasch.org/rmt/rmt122m.htm>
- Du, Y. (1995) When to adjust for Differential Item Functioning. *Rasch Measurement Transactions*, 9(1), 414.
- Elliott, R., Fox, C. M., Belyukova, S. A., Stone, G. E., Gunderson, J., & Zhang, X. (2006). Deconstructing therapy outcome measurement with Rasch analysis: The SCL-90-R. *Psychological Assessment*, 18, 359-372.
- Embretson, S. (2006). The continued search for nonarbitrary metrics in psychology. *American Psychologist*, 61, 50-55.
- Eysenck, M. W., & Keane, M. T. (2007). Percepção da leitura e da fala. In M. W. Eysenck & M. T. Keane (Eds.), *Manual de Psicologia Cognitiva* (pp. 313-347). Porto Alegre, RS: Artmed.
- Gilbert, R., Martínez, G., & Vidal-Abarca, E. (2005). Some good text are always better text revision to foster inferences of readers with high and low prior background knowledge. *Learning and Instruction*, 15, 45-68.
- Guidetti, A. A., & Martinelli, S. C. (2007). Compreensão em leitura e desempenho em escrita de crianças do ensino fundamental. *PSIC - Revista de Psicologia da Vetor Editora*, 8(2), 175-184.
- Joly, M. C. R. A. (2009). Estudos com o sistema orientado de cloze para o ensino fundamental. In A. A. A Santos, E. Boruchovitch, & K. L. Oliveira (Eds.), *Cloze: Um instrumento de diagnóstico e intervenção* (pp. 119-145). São Paulo, SP: Casa do Psicólogo.
- Kintsch, W. (1994). Text comprehension, memory and learning. *American Psychologist*, 49, 294-303.
- Kintsch, W., & Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363-394.
- Linacre, J. M., & Wright, B. D. (2004). Construction of measures from many-facet data. In E. V. Smith Jr., & R. M. Smith (Eds.), *Introduction to Rasch measurement: Theory, models, and applications* (pp. 296-321). Maple Grove, MN: JAM.
- Linacre, J. M., & Wright, B. D. (2009). *WINSTEPS: Multiple-choice, rating scale, and partial credit Rasch analysis* [Computer software]. Chicago, IL: MESA Press.
- Lopes, D. M. B., & Paula, E. M. (2008). Habilidades de resolução de conflito e ocorrência de disfluências comuns em crianças em desenvolvimento normal de linguagem. *Revista da Sociedade Brasileira de Fonoaudiologia*, 13(3), 272-278.
- Muth, K. D. (1989). *Children's comprehension of text*. Newark, DE: International Reading Association.
- Myford, C. M., & Wolfe, E. W. (2004). Detecting and measuring rater effects using many-facet Rasch measurement: Part II. *Journal of Applied Measurement*, 5, 189-227.
- Oliveira, K. L., Boruchovitch, E., & Santos, A. A. A. (2009). A técnica de cloze na avaliação da compreensão em leitura. In A. A. A. Santos, E. Boruchovitch, & K. L. Oliveira (Eds.), *Cloze: Um instrumento de diagnóstico e intervenção* (pp. 47-77). São Paulo, SP: Casa do Psicólogo.
- Olson, M. W. (1990). *Opening the door to classroom research*. Newark, DE: International Reading Association.
- Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, 20(1), 26-56.
- Primi, R., Carvalho, L. F., Miguel, F. K., & Silva, M. C. R. (2010). Análise do funcionamento diferencial dos itens dos exame nacional do estudante (CENADE) de Psicologia de 2006. *Psico-USF*, 15(3), 379-393.
- Rupp, A. A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing*, 23(4), 441-474.
- Salles, J. F., & Parente, M. A. M. P. (2007). Avaliação da leitura e escrita de palavras em crianças de 2ª série: Abordagem neuropsicológica cognitiva. *Psicologia: Reflexão e Crítica*, 20(2), 220-228.
- Santos, A. A. A. (2004). O Cloze como técnica de diagnóstico e remediação da compreensão em leitura. *Interação em Psicologia*, 8(2), 217-226.
- Santos, A. A. A. (2005). *Evidências de validade de critério para o Teste de Cloze*. Manuscrito não publicado, Programa de Pós-Graduação *Stricto-Sensu* em Psicologia, Universidade São Francisco, Itatiba, SP.
- Santos, A. A. A., Carvalho, L. F., Monteiro, R. M., & Alcará, A. R. (2010). *Uso da TRI para uma medida de avaliação da compreensão em leitura*. Trabalho apresentado no Seminário Internacional “Contributos da Psicologia em Contextos Educativos”, Instituto de Educação, Universidade do Minho, Portugal.
- Santos, A. A. A., & Oliveira, E. Z. (2010). Avaliação e desenvolvimento da compreensão em leitura no ensino fundamental. *Psico-USF*, 15(1), 93-102.
- Santos, A. A. A., Primi, R., Taxa, F. O. S., & Vendramini, C. M. M. (2002). O teste de cloze na avaliação da compreensão em leitura. *Psicologia: Reflexão e Crítica*, 15(3), 549-560.
- Singer, H., & Donlan, D. (1982). Problem-solving schema with question generation for comprehension of complex short stories. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (pp. 166-186). Newark, DE: IRA.
- Smith, E. V., Jr. (2001). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement*, 2(3), 281-311.
- Spinillo, A. G., & Mahon, E. R. (2007). Compreensão de texto em crianças: Comparações entre diferentes classes de inferência a partir de uma metodologia on-line. *Psicologia: Reflexão e Crítica*, 20(3), 463-471.

Carvalho, L. F., Monteiro, R. M., Alcará, A. R. & Santos, A. A. A. (2013). Aplicação da TRI em uma Medida de Avaliação da Compreensão de Leitura.

- Sternberg, R. J. (2008). Linguagem em contexto. In R. J. Sternberg (Ed.), *Psicologia cognitiva* (pp. 326-364). Porto Alegre, RS: Artmed.
- Taylor, W. L. (1953). Close procedure: A new tool for measuring readability. *Journalism Quarterly*, 30, 415-433.
- Tennant, A., & Pallant, J. F. (2007). DIF matters: A practical approach to test if Differential Item Functioning makes a difference. *Rasch Measurement Transactions*, 20(4), 1082-1084.
- Urbina, S. (2007). *Fundamentos da testagem psicológica*. Porto Alegre, RS: Artmed.
- Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2008). Comparability of Computer-based and Paper-and-Pencil Testing in K-12 Reading Assessments: A meta-analysis of testing mode effects. *Educational and Psychological Measurement*, 68(1), 5-24.
- Witter, G. (2010). Motivação e leitura. In E. Boruchovitch, J. A. Bzuneck, & S. E. R. Guimarães (Eds.). *Motivação para aprender: Aplicações no contexto educativo* (pp. 169-192). Petrópolis, RJ: Vozes.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA.

Recebido: 05/05/2011
1ª revisão: 19/12/2011
Aceite final: 26/12/2011