



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE ESTATÍSTICA



Comparação da Teoria Clássica dos Testes com a Teoria de Resposta ao Item aplicado ao subteste Semelhanças da WASI

Autor: Fernanda Maciel de Quadros
Orientador: Professora Dr^a. Suzi Alves Camey

Porto Alegre, 05 de Julho de 2010.

Universidade Federal do Rio Grande do Sul
Instituto de Matemática
Departamento de Estatística

Comparação da Teoria Clássica dos Testes com a Teoria de Resposta ao Item aplicado ao subteste Semelhanças da WASI

Autor: Fernanda Maciel de Quadros

Monografia apresentada para obtenção
do grau de Bacharel em Estatística.

Banca Examinadora:
Professora Dr.^a. Suzi Alves Camey (orientadora)
Professora Dr.^a. Clarissa Marcelli Trentini

Porto Alegre, 05 de Julho de 2010.

Este artigo será submetido a “Psicologia: Reflexão & Crítica”

Comparação da Teoria Clássica dos Testes com a Teoria de Resposta ao Item aplicado ao subtteste Semelhanças da WASI

Fernanda Maciel de Quadros

Clarissa Marcelli Trentini

Denise Balem Yates

Suzi Alves Camey

Resumo

O presente artigo é resultante de uma pesquisa que teve por objetivo analisar os itens do subtteste Semelhanças da Escala de Inteligência Wechsler Abreviada (WASI). Os itens foram analisados através da Teoria Clássica dos Testes (TCT) e da Teoria de Resposta ao Item (TRI) a partir das quais foram obtidos a ordenação e classificação dos mesmos de acordo com o seu parâmetro de dificuldade. Também foram estabelecidos os critérios de início e fim do teste. Na TRI foi utilizado o modelo de Rasch. Nesta aplicação não foram encontradas diferenças entre os dois métodos no que diz respeito ao ordenamento e classificação do grau de dificuldade. Pelos dois métodos foi observada a ausência de índices com maior dificuldade.

Palavras-chave: Teoria Clássica dos Testes (TCT); Teoria de Resposta ao Item (TRI); Escala de Inteligência Wechsler Abreviada (WASI) .

Abstract

The aim of this paper is to examine items in the subttest similarities of the Wechsler Abbreviated Scale of Intelligence (WASI). The sorting and classification by difficulty degree were obtained from the items analyze using the Classical Test Theory (CTT) and Item Response Theory (IRT). Criterias to begin and end the test were set. The Rasch

Model was the employed at IRT. The results of both tests – IRT and CTT- shows no differences in the sorting and classification. Indexes with greater difficulty are absent.

Keywords: Classical Test Theory (CTT); Item Response Theory (IRT); Wechsler Abbreviated Scale of Intelligence (WASI).

A partir do século XIX tem-se registros da preocupação em se medir fenômenos psicológicos. Filósofos, psicólogos e estatísticos, entre outros, preocupavam-se em encontrar um meio de mensurar traços latentes como sensações, emoções e comportamentos humanos (Chachamovich, 2007). De acordo com Pasquali (2003), tal medição deu origem a ciência denominada psicometria, cuja origem encontra-se nos trabalhos do estatístico Spearman, que no que se refere a psicologia, inspirou-se no estudioso Francis Galton. Na década de 1880, Galton iniciou seus estudos influenciando os trabalhos tanto na área da psicometria prática quanto na teórica. Em 1890, foi a vez de Cattell desenvolver medidas de diferenças individuais e apresentar a terminologia de teste mental. Na década de 1900, Spearman lançou os fundamentos da teoria clássica, essa é considerada por Pasquali (2003), a década de Spearman. Nesta época também destaca-se o trabalho de Binet que visava a avaliação das aptidões humanas.

A psicometria surge através de estatísticos, mas segundo Pasquali (2003) ela é uma área da psicologia com uma concepção estatística, que explica os comportamentos e aptidões por meio de testes cuja mensuração é feita através das respostas que os indivíduos fornecem aos itens. Neste sentido, duas são as teorias para tal fim. A Teoria Clássica dos Testes (TCT) tem por objetivo a interpretação da resposta final, ou seja, o que a soma dos itens diz sobre o sujeito. Já a Teoria de Resposta ao Item (TRI) tem o propósito de medir a habilidade do sujeito de acordo com as respostas dadas a cada item. Ou seja, a TCT analisa o resultado final e a TRI analisa partes e probabilidades que geram o resultado final (Pasquali, 2003).

Independente da teoria utilizada, a respeito da distribuição dos itens de acordo com as suas dificuldades, Pasquali (2003) sugere que itens muito fáceis ou muito difíceis não trazem nenhum tipo de informação, ele exemplifica dizendo que indivíduos que acertam 50% das questões e erram 50% das questões fornecem 2.500 (50x50) comparações possíveis e os que erram ou acertam todas, ou seja, 0% e 100%, fornecem 0 de informação (0x100). Ele propõe que se os itens que oferecem maior informação são os que têm uma probabilidade 0,50 de acerto, então os itens deveriam estar mais ou menos distribuídos dentro de uma curva normal. Ou seja, a distribuição dos itens de acordo com as suas dificuldades seria de tal maneira que 10% dos itens ficassem em cada uma das duas faixas extremas (itens muito fáceis ou muito difíceis), 20% em cada uma das faixas seguintes (itens fáceis ou difíceis) e 40% na faixa média (itens medianos).

A preocupação com o ordenamento de itens se deve ao fato de que, segundo Figueiredo, Mattos, Pasquali & Freire (2008), todos os testes devem incluir itens muito fáceis para motivar os sujeitos e aumentar o seu grau de cooperação no teste. Os itens iniciais devem ser relativamente fáceis e devem avançar para itens mais difíceis de forma crescente permitindo uma maior confiança dos sujeitos e reduzindo a probabilidade de perderem tempo com itens que tenham um grau de dificuldade maior que suas habilidades (Anastasi & Urbina, 2000).

Neste trabalho foram analisados itens do subteste Semelhanças do instrumento WASI (Wechsler Abbreviated Scale of Intelligence/Escala de Inteligência Wechsler Abreviada) em que as referidas informações foram submetidas à análise pela Teoria Clássica dos Testes (TCT) e pela Teoria de Resposta ao Item (TRI).

Teoria Clássica dos Testes (TCT)

Na TCT as aptidões são medidas, em geral, pela soma dos itens. De acordo com Pasquali (2003), uma grande preocupação dos psicometristas era com o erro contido nessa soma. Quando o sujeito responde uma questão, ele recebe uma pontuação por essa

resposta, no final da aplicação do teste teremos um escore final, que é a soma dos pontos. No entanto, o autor alerta para o fato de que este escore representa uma magnitude, seja depressão, inteligência, no caso deste estudo, a capacidade do indivíduo em diferenciar palavras semelhantes. Neste resultado está contido um erro, pois trata-se de uma operação empírica e que está sujeita a erros. Deste modo, conclui-se que o escore bruto é formado pelo escore total menos um erro. A dificuldade do item na TCT é definida em termos de percentagens de acertos, quanto mais próximo de 100% a taxa de acertos mais fácil é o item (Pasquali, 2003).

No presente estudo serão analisados os escores totais dos 26 itens que compõem o subteste Semelhanças, outro ponto de interesse é avaliar se existem pontos de início e fim da avaliação e o ordenamento dos itens de acordo com o seu grau de dificuldade. A pesquisa medirá a capacidade do indivíduo em diferenciar figuras e palavras semelhantes, recebendo uma pontuação estabelecida pelo instrumento. O ponto de início determinado na TCT é o critério adotado pelas Escalas Wechsler completas adaptadas para o Brasil, tanto na WISC-III quanto na WAIS-III, foi utilizada a percentagem de pelo menos 95% de acertos nos itens. O critério de parada é o número de erros consecutivos para o qual a probabilidade de acerto do próximo item é menor do que 0,05 (Nascimento & Figueiredo, 2002).

Teoria de Resposta ao Item (TRI)

A Teoria da Resposta ao Item (TRI), teoria da psicometria moderna, é utilizada para medir os traços latentes, traços estes que são características intrínsecas dos indivíduos que não podem ser medidas diretamente. A referida teoria trabalha com a probabilidade de um indivíduo acertar um item de acordo com sua proficiência ou traço latente (Andrade, Tavares & Valle, 2000). Neste sentido, segundo Pasquali & Primi (2003), a TRI teria uma relação de causa-efeito sendo o traço latente a causa e a resposta do indivíduo o efeito, cujos estímulos para que esta relação ocorresse seriam os itens do teste. Alguns dos

questionamentos que são respondidos através da TRI, foram feitos no início do século XX, mas a teoria só começou a se desenvolver, de fato, nos anos 80, após o avanço da informática, devido a sua complexidade matemática (Pasquali, 2003).

Vários são os modelos logísticos para calcular o escore TRI, cuja diferenciação diz respeito à quantidade de parâmetros e o tipo de resposta dada ao item. Neste trabalho, o foco são os itens com repostas dicotômicas (por exemplo, certo/errado). Neste contexto, três modelos são os mais conhecidos: (i) o modelo logístico de um parâmetro que trabalha com a dificuldade do item; (ii) o modelo logístico de dois parâmetros que se baseia na dificuldade e na discriminação do item e; (iii) o modelo de três parâmetros que considera a dificuldade, a discriminação do item e o acerto ao acaso (Andriola, 2009).

No presente estudo, a análise dos dados será baseada no modelo logístico de um parâmetro, também conhecido como modelo de Rasch. Modelo este desenvolvido por volta do ano de 1960, por George Rasch (Andriola, 2009), que é apenas influenciado pela dificuldade do item:

$$P(X_i = 1 | \theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}} \quad (1)$$

onde:

- θ é a variável latente, ou seja, a proficiência do indivíduo medida por aquele item;
- $P(X_i = 1 | \theta)$ é a probabilidade de acerto do item i dado θ ;
- b é o parâmetro de dificuldade do item;
- D é uma constante;
- e é uma constante com valor aproximado de 2,72.

Importa assinalar que nos casos em que a constante D assumir o valor de 1,7, a função logística torna-se igual à normal acumulada (ogiva).

De acordo, com Andriola (2009) a probabilidade de um indivíduo acertar um determinado item, dependerá da sua proficiência no assunto, ou seja, da medida de θ e da dificuldade do item, embora o autor cite outros parâmetros, será considerado apenas a dificuldade do item para este trabalho. O parâmetro de dificuldade do item é, segundo Pasquali (2003), o ponto na escala de habilidade onde a probabilidade de uma resposta correta é 0,50. A dificuldade é definida em termos do traço latente, ela é diretamente proporcional ao nível da habilidade do sujeito. Quanto maior a dificuldade do item, maior aptidão o indivíduo que responde ao teste deverá apresentar.

No caso da relação funcional acima referida (1), esta pode ser representada graficamente através da curva característica do item (CCI).

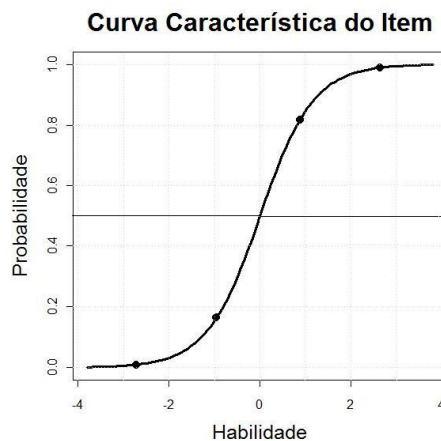


Figura 1. Curva Característica do Item

Na Figura 1, vislumbra-se a CCI, que é considerada por Pasquali (1997) a identidade do item. Observamos na figura, que quanto menor a habilidade, menor a probabilidade de acerto. Na Figura 1, indivíduos com habilidades próximas a zero (ou seja, aptidões medianas) tem probabilidade de 0,50 de acertar o item.

As análises do presente estudo serão feitas através da função de um parâmetro (1), definida anteriormente, que será utilizada para a definição dos critérios de início e

ordenamento dos itens, com o propósito de comparar os resultados obtidos através da TCT.

Instrumento WASI (Escala de Inteligência Wechsler Abreviada)

Este estudo como já referido anteriormente, está baseado no subteste Semelhanças do instrumento WASI. Trata-se de um instrumento que mede a capacidade intelectual do indivíduo, que deve ser administrado individualmente. Para tanto, a aplicação tem a duração média de trinta minutos e é indicada para indivíduos que tenham entre 6 (seis) a 89 (oitenta e nove) anos (Yates, Trentini, Tosi, Corrêa, Poggere, & Valli, 2006).

Este instrumento é composto por quatro subtestes: dois verbais e dois de execução. Os verbais são divididos em Semelhanças e Vocabulário e os de execução em Cubos e Raciocínio Matricial. O subteste Semelhanças é composto por 26 itens, sendo os 4 primeiros de combinação de figuras e os 22 restantes de conceitos semelhantes. Ou seja, ele mede a capacidade verbal dos indivíduos em diferenciar figuras e palavras semelhantes.

Métodos

Participantes

Os sujeitos avaliados neste trabalho compõem uma subamostra oriunda do trabalho “Explorando as Propriedades Psicométricas da Escala de Inteligência Wechsler Abreviada” e é composta por 155 pessoas. A faixa etária selecionada varia de 20 a 64 anos, sendo 74 (47,7%) do sexo masculino e 81 (52,3%) do sexo feminino. Em relação à escolaridade, 55 (35,5%) dos sujeitos tem entre 1 e 7 anos de estudo, 23 (14,8%) entre 8 e 11 anos, 77 (49,7%) mais de 11 anos. Dos respondentes 92 (62,2%) nunca tiveram repetências. A amostra utilizada apresenta indivíduos com QI maior que 70, segundo as Matrizes Progressivas de Raven, para tanto foram eliminados 13 indivíduos da amostra inicial que era composta por 168 sujeitos.

Instrumento

Utilizou-se a adaptação para o português do subteste Semelhanças do Instrumento WASI (Wechsler Abbreviated Scale of Intelligence). A pontuação dos itens foi feita da seguinte forma: 0 para respostas incorretas, 1 para respostas corretas simples e 2 para respostas corretas elaboradas. Neste trabalho, as respostas foram dicotomizadas, ou seja, 0 para respostas incorretas e 1 para respostas corretas (pontuação 1 ou 2).

Procedimentos

Os participantes residem na cidade de Porto Alegre, o recrutamento foi realizado junto a escolas, universidades e empresas. Todas as questões foram aplicadas, independente da idade. Foi preenchido o Termo de Consentimento Livre e Esclarecido e uma ficha de dados para a coleta de informações sociodemográficas. Para os maiores de 60 anos foi feita uma triagem para identificação e exclusão de casos de demência com o Mini-Exame do Estado Mental.

Resultados

Os itens foram ordenados de acordo com a TCT e a TRI. Em ambas, utilizou-se o parâmetro da dificuldade, os resultados foram calculados através do pacote ltm, no software R (Rizopoulos, 2006). Entretanto, na TCT foi calculado o percentual de acertos em cada item e ordenado de forma decrescente. Já na TRI, o ordenamento foi feito pelo parâmetro de dificuldade exposto na ordem crescente, conforme a Tabela 1.

Tabela 1. Dificuldade do item

Itens	Dificuldade do item - TCT (%)	Dificuldade do item - TRI	$P(X_i=1 \theta=0)$
13	98,06	-3,21839	0,9958385
3	98,06	-3,21826	0,9958376
1	98,06	-3,21778	0,9958341
2	95,48	-2,63433	0,9888337
8	95,48	-2,63403	0,9888280
6	94,19	-2,44533	0,9846617
7	93,55	-2,36368	0,9824154
4	93,55	-2,36367	0,9824149
9	93,55	-2,36366	0,9824147
5	91,61	-2,15066	0,9749236
10	90,32	-2,02829	0,9692951
14	89,68	-1,97119	0,9662672
17	81,94	-1,42321	0,9185117
16	80,00	-1,31053	0,9029564
11	73,55	-0,97436	0,8400189
12	69,68	-0,79348	0,7942105
18	64,52	-0,56911	0,7248438
15	58,71	-0,33307	0,6380440
22	57,42	-0,28249	0,6179369
23	56,77	-0,25729	0,6077583
20	55,48	-0,20704	0,5871965
21	54,84	-0,18241	0,5769979
19	52,90	-0,10844	0,5460124
24	51,61	-0,05959	0,5253334
26	30,97	0,72464	0,2255971
25	26,45	0,91295	0,1745313

Apesar do ordenamento dos itens ser o mesmo na TCT e na TRI, como pode ser observado na tabela 1, pode-se apontar uma pequena diferença no parâmetro de dificuldade da TRI nos itens 13-3-1, 2-8 e 7-4-9, que pela TCT apresentavam o mesmo percentual de acerto. Esta é uma das vantagens da TRI, pois através dela consegue-se estabelecer uma ordem de dificuldade, o que não aconteceria se fosse utilizada somente a TCT.

O critério de início na TCT é definido de acordo com as Escalas Wechsler completas, tanto na WISC-III quanto na WAIS, e é utilizado o percentual de acerto de pelo menos 95%, como já mencionado anteriormente. De acordo com esta definição, inicia-se o teste no item 2 ou 8, porque são os itens mais difíceis entre aqueles com percentual de acertos maior do que 95%.

Como critério de início, pela TRI, sugere-se que seja utilizada a probabilidade de que uma pessoa com habilidade mediana acerte ao item, $P(X_i = 1 | \theta = 0)$. Não há definição para uma probabilidade mínima de acerto. Conforme análise da Tabela 1, se

fosse utilizado 0,98 como ponto de corte para essa probabilidade, o teste se iniciaria no item 9 e não no item 8. Por outro lado, se definirmos o ponto de corte como 0,99, então começaríamos no item 1, pulando os dois itens iniciais. Neste caso, a maioria das pessoas, provavelmente até as com baixas proficiências para o assunto acertariam todos os itens iniciais.

Na Tabela 2 são apresentadas as probabilidades de erros e acertos, depois de um determinado número de erros consecutivos. Essas probabilidades serão utilizadas para definição do critério de parada (Nascimento & Figueiredo, 2002). No presente estudo, este critério ficou estabelecido em 7 erros consecutivos, pois neste caso a probabilidade de que um indivíduo acerte uma questão depois de apresentar 7 erros sucessivos é de 5%.

Tabela 2. Número de erros consecutivos

	Número de erros consecutivos						
	1	2	3	4	5	6	7
Erro	0,74	0,84	0,89	0,92	0,93	0,94	0,95
Acerto	0,26	0,16	0,11	0,08	0,07	0,06	0,05

Os itens foram classificados como muito fáceis, fáceis, medianos, difíceis e muito difíceis, segundo o critério sugerido por Pasquali (2003), citado na introdução. A Figura 2 ilustra a distribuição dos itens nas 5 categorias e compara as duas teorias (TCT e TRI) com os valores esperados para tais níveis de dificuldade. Percebe-se, principalmente nas faixas extremas, que os resultados observados ficaram longe do esperado.

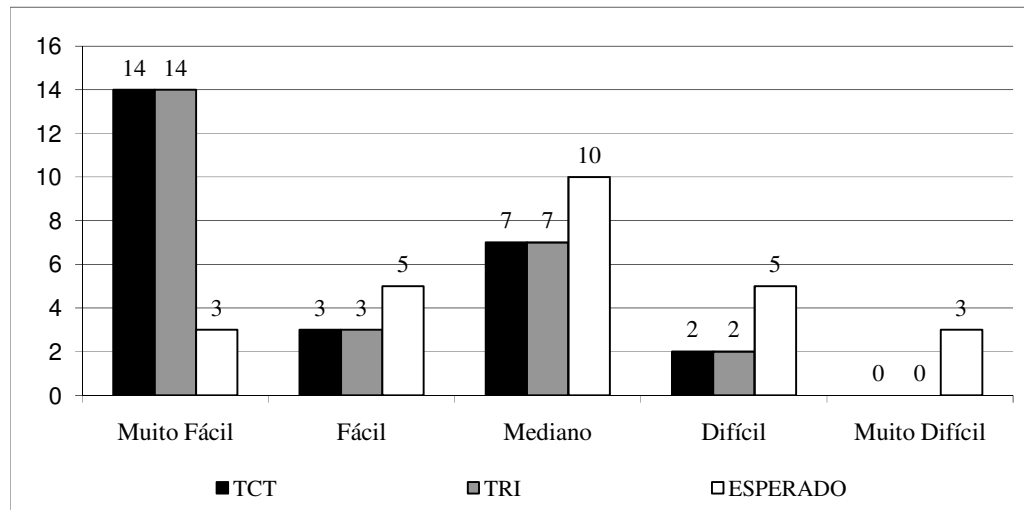


Figura 2. Distribuição dos itens nas 5 categorias e valores esperados

Pode-se ver na Figura 3, a curva característica do item que mostra relação entre a habilidade dos sujeitos e a probabilidade de acertos para alguns itens. Os itens abaixo representados foram escolhidos para ilustrar a separação em categorias de dificuldade, o item 13 como pode ser observado na tabela 1 foi classificado como item muito fácil, o item 11 foi considerado um item fácil, o 15 um item mediano e o item 26, difícil. Uma análise minuciosa permite ver claramente esta divisão de dificuldade, que pode ser observada através do ponto de inflexão das curvas, ou seja, quanto mais rápido ela atinge a probabilidade 1, mais fácil é o item. Assim, vislumbra-se que os itens atingem rapidamente este nível de probabilidade de acerto, o que reflete o que havia sido concluído anteriormente acerca da ausência de itens muito difíceis e a presença de itens muito fáceis.

Curva Característica do Item

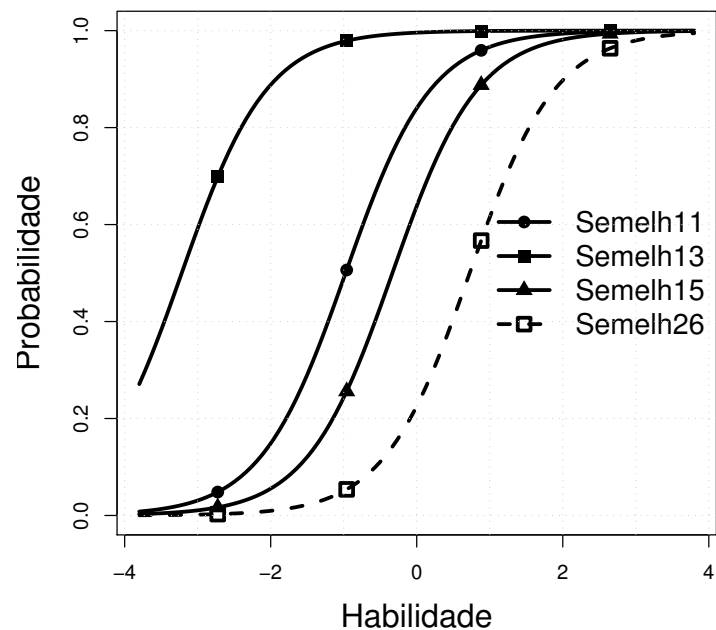


Figura 3. CCI para os itens 11, 13, 15 e 26

Considerações Finais

De acordo com a corrente doutrinária, embora a TRI não substitua integralmente a TCT, ela traz grandes contribuições e substitui apenas parte da Teoria Clássica. No presente estudo, isso ficou demonstrado, porque embora o ordenamento dos itens não tenha diferido em relação às duas teorias, segundo o modelo de um parâmetro, a TRI demonstrou a vantagem de estabelecer uma ordem de dificuldade para os itens. Um problema encontrado foi a falta de itens muito difíceis e poucos itens difíceis, pois desta forma mesmo indivíduos com habilidades baixas ou medianas podem apresentar altos escores. Isso é mais problemático na TCT, onde o escore é baseado simplesmente na soma dos acertos, sem considerar a dificuldade do item.

Quanto ao critério de início do teste, como ainda não há critério análogo à TCT na TRI, não é possível comparar as duas teorias. Um fato a ser destacado é que no caso de empate entre os percentuais de acertos, não há um critério objetivo para determinar onde o

teste deve iniciar. Para a utilização da TRI na definição de um ponto de início a probabilidade de uma pessoa com habilidade mediana acertar um item, fica sugerido. Propõe-se uma discussão sobre dois pontos de corte: a probabilidade de acerto de pelo menos 0,98 ou 0,99. Para os dois pontos sugeridos, o item de início seria diferente do ponto de início estabelecido pela TCT.

Para a amostra estudada, como reflexo do nível de dificuldade dos itens, são necessários muitos erros consecutivos (sete) para que o teste seja interrompido.

Uma limitação deste artigo é não explorar outros modelos de TRI, incluindo aqueles para respostas politômicas e modelos com mais parâmetros.

Estudos recentes que relacionam a TCT com a TRI demonstram a preocupação de uniformidade nos testes, ou seja, testes que consigam discriminar as pessoas de acordo com seus níveis de traço latente. Para isso, o teste necessita desde itens muito fáceis até itens muito difíceis, e é neste ponto que a TRI tem se destacado em relação à TCT, uma vez que uma análise mais detalhada é possível através da TRI e a dificuldade dos itens pode ser colocada na mesma escala da distribuição de habilidade dos indivíduos. Castro (2008) salienta que modelos TRI tem grande vantagem no sentido em que fornecem estimativas para os itens e para o traço latente e não variam com as características da população com respeito ao traço latente, ou seja, os resultados independem do instrumento avaliado e dos indivíduos selecionados.

Referências

Andrade, D. F., Tavares, H. R. & Valle, R. C. Teoria de Resposta ao Item: conceitos e aplicações. ABE — Associação Brasileira de Estatística, 4º SINAPE, 2000.

Andriola, W.B. (2009). Psicometria Moderna: Características e Tendências. Estudos em Avaliação Educacional, v.20, n.43, 319-340.

Anastasi, A. & Urbina, S. (2000). *Testagem Psicológica*. Porto Alegre: Artes Médicas.

Castro, S.M.J. 2008. *Teoria da Resposta ao Item: Aplicação na Avaliação da Intensidade de Sintomas Depressivos*. Unpublished doctoral dissertation, Programa de Pós-Graduação em Epidemiologia, Faculdade de Medicina, Universidade Federal do Rio Grande do Sul. Porto Alegre, Brasil.

Chachamovich, E. (2007). *Teoria de resposta ao item: aplicação do modelo Rasch em desenvolvimento e validação de instrumentos em saúde mental*. Unpublished doctoral dissertation, Programa de Pós-Graduação em Ciências Médicas: Psiquiatria, Faculdade de Medicina, Universidade Federal do Rio Grande do Sul. Porto Alegre, Brasil.

Figueiredo, V.L.M, Mattos, V.L.D, Pasquali, L. & Freire, A.P. (2008). *Propriedades Psicométricas dos Itens do Teste WISC-III*. *Psicologia em estudo*, Maringá, v.13, n.3, p. 585-592.

Nascimento, E. & Figueiredo, V.L.M.(2002). *WISC-III e WAIS-III: Alterações nas Versões Originais Americanas Decorrentes das Adaptações para o Uso no Brasil*. *Psicologia: Reflexão e Crítica*, 15(3), 603-612.

Pasquali, L. (1997). *Psicometria: teoria e aplicações*. Brasília: Editora UnB.

Pasquali, L. (2003). *Psicometria: teoria dos testes na psicologia e na educação*. Petrópolis, RJ: Vozes.

Pasquali, L., & Primi, R. (2003). *Fundamentos da Teoria da Resposta ao Item – TRI*. *Avaliação Psicológica*, 2, 99-110.

R Development Core Team (2009). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

Rizopoulos, D. 2006. *Ltm: An R Package for Latent Variable Modeling and Item Response Theory Analyses*. *Journal of Statistical Software*, v.17.

Yates, D. B., Trentini, C. M., Tosi, S. D., Corrêa, S. K., Poggere, L.C., & Valli, F. (2006). Apresentação da Escala de Inteligência Wechsler Abreviada (WASI). *Avaliação Psicológica*, 5(2), 227-233.