

## **Análise do impacto da dificuldade dos cadernos na proficiência dos alunos através do uso da Modelagem Rasch Multifacetada**

Wellington Silva<sup>1</sup>, Neimar S Fernandes<sup>2</sup>, Joaquim S Neto<sup>3</sup>, Alicia Bonamino<sup>4</sup>

<sup>1</sup> CAEd/UFJF, wellington@caed.ufjf.br

<sup>2</sup> CAEd/UFJF, neimar@caed.ufjf.br

<sup>3</sup> UNB, jjsoaresneto@gmail.com

<sup>4</sup> PUC-Rio, aliciamcbonamino@gmail.com

### **Resumo**

Na perspectiva de obter resultados mais equânimes ao nível dos alunos, neste trabalho, é utilizado o modelo Rasch multifacetado para minimizar os efeitos da utilização de diferentes modelos de cadernos de testes para diferentes alunos.

**Palavras-chave:** Modelos Rasch Multifacetados, Efeito do *Design* do teste, Estimação de Proficiências

### **1. Introdução**

As avaliações em larga escala são uma riquíssima fonte de aplicação da TRI e ao mesmo tempo em que levantam questões muito pertinentes aos seus pressupostos.

É sabido que o foco de uma avaliação em Larga Escala é a escola, no máximo a turma, portanto é tolerável se suportar um grau maior de imprecisão na medida do aluno, em troca de uma melhor interpretação pedagógica dos resultados, através de uma maior abrangência tanto nos descritores quanto nos diferentes níveis de dificuldade dos itens. Nesse sentido, os alunos são submetidos a diferentes modelos de cadernos, através da utilização de Blocos Incompletos Balanceados – BIB e a produção de uma medida de conhecimento através da utilização de modelos estatísticos pela Teoria da Resposta ao Item que equaliza todos os diferentes cadernos em uma mesma escala. Normalmente, no Brasil, utiliza-se a modelagem com apenas duas facetas, ou seja, dificuldade do item e proficiência do aluno.

Porém, embora tenha-se a preocupação de equilibrar a dificuldade dos diferentes modelos de cadernos na avaliação, esse balanceamento não é perfeito. O que se observa na prática, é a influência da dificuldade dos cadernos no desempenho dos alunos. Nesse artigo, analisaremos esse efeito através da utilização da modelagem: Rasch multifacetada - *Many-Facet Rasch Measurement* – MFRM (Eckes, 2005), na qual as características dos cadernos de testes entrarão como facetas adicionais na modelagem pela TRI.

## 2. Metodologia

Os dados utilizados no presente trabalho são oriundos de uma amostra de 10 000 (dez mil) alunos do 5º ano da Rede Pública do estado de Minas Gerais, em 2016, utilizou-se uma base de respostas dos alunos, autorizada pela Secretaria de Estado de Educação, com identificadores mascarados e sorteada por amostragem aleatória simples.

O *design* do teste utilizado é formado por 21 modelos de cadernos, sendo cada caderno composto por duas disciplinas, a saber: Língua Portuguesa (P) e Matemática (M). Para montagem dos itens nos cadernos utilizou-se a técnica do BIB, conforme apresentada no Quadro I.

**Quadro I:** *Design* do teste e proficiência em Língua Portuguesa para alunos submetidos a diferentes modelos de cadernos.

Caderno	POSIÇÃO				Proficiência		Caderno	POSIÇÃO				Proficiência	
	1	2	3	4	Média	Des. Pad		1	2	3	4	Média	Des. Pad
1	P01	P02	M01	M02	0,033	0,989	2	M02	M03	P02	P03	-0,099	1,027
3	P03	P04	M03	M04	0,089	0,989	4	M04	M05	P04	P05	-0,126	1,022
5	P05	P06	M05	M06	0,146	0,971	6	M06	M07	P06	P07	-0,035	1,022
7	P07	P01	M07	M01	0,102	0,971	8	M01	M03	P01	P03	-0,116	1,044
9	P02	P04	M02	M04	-0,020	0,987	10	M03	M05	P03	P05	-0,079	1,003
11	P04	P06	M04	M06	0,079	0,987	12	M05	M07	P05	P07	-0,014	0,955
13	P06	P01	M06	M01	0,069	0,979	14	M07	M02	P07	P02	-0,139	1,022
15	P01	P04	M01	M04	0,141	1,034	16	M02	M05	P02	P05	-0,108	0,987
17	P03	P06	M03	M06	0,066	1,002	18	M04	M07	P04	P07	-0,013	1,010
19	P05	P01	M05	M01	0,085	0,971	20	M06	M02	P06	P02	-0,102	0,993
21	P07	P03	M07	M03	0,024	0,965	Todos os cadernos					0	1

Fonte: CAEd/UFJF, 2016

Nesse *design*, cada bloco é composto por 11 itens, portanto cada aluno responde a 22 itens de Língua Portuguesa e a 22 itens de Matemática. Onze testes iniciam-se por Língua Portuguesa e dez, por Matemática. Os blocos alternam entre os cadernos, tentando-se ao máximo equilibrar o número de vezes que aparecem no início e ao final do caderno.

A fim de se analisar o modelo usualmente aplicado, utilizou-se o modelo da TRI a 3 parâmetros (3PL) para um único grupo, sem fixar nenhum item, nem utilizar a proficiência oficial na calibração dos parâmetros, para que, assim, fosse evidenciada, ao máximo, a relação entre padrão de respostas e desempenho. O pacote utilizado foi o MIRT (CHALMERS, 2012) do *Software R*.

Através desse modelo, podemos constatar, no Quadro 1, 21 subgrupos da população obtendo médias distintas de proficiência, em função do modelo de caderno que realizaram. Podemos verificar, por exemplo, que a diferença de proficiência entre os alunos que fizeram o caderno 14 (-0,139) e 5 (0,146) é de 0.2854, ou seja, a diferença observada chega a mais que um quinto de desvio padrão, em duas subpopulações que deveriam possuir a mesma média e desvio padrão, admitindo apenas variações amostrais.

Nota-se que todos os cadernos pares (Língua Portuguesa no final do caderno) se situam abaixo da média, já entre os ímpares (Língua Portuguesa no início do caderno) tal fenômeno ocorre em apenas um caderno. Também é possível constatar que 7 dos dez cadernos pares possuem desvio padrão superior ao desvio padrão geral, enquanto entre os ímpares apenas 2 em 11 cadernos manifestam tal fenômeno.

Em nossos estudos, realizados apenas para Língua Portuguesa, consideraremos as características dos diferentes cadernos de testes como outras facetas na modelagem via MFRM, no intuito de identificar quais dessas facetas são mais significativas para um melhor equilíbrio entre as proficiências dos alunos. Pois, conforme apresentado anteriormente, através dos resultados obtidos pela modelagem via 3PL, detectamos diferenças significativas entre os mesmos.

### Modelagem MFRM

O modelo de 3 Parâmetros se torna extremamente complexo no caso de se considerar as inúmeras circunstâncias em torno do item, portanto, optou-se por um modelo mais simples, o de Modelo de *Rasch*, de forma a se poder trabalhar com mais facetas, conforme apresentado no modelo a seguir:

$$P(Y_{ij} = 1 | \theta_i, \xi_j) = \text{logit}(\theta_i - \sum_{k=1}^F b_{jk})$$

$K=1, 2 \dots F$  sendo  $F$  número total de facetas utilizadas

$Y_{ij}$  resposta do  $i$ -ésimo aluno ao  $j$ -ésimo item

$\theta_i$  traço latente do  $i$ -ésimo aluno

$b_{jk}$  dificuldade da  $k$ -ésima faceta para o  $j$ -ésimo item

Nota-se que, nesse modelo, por simplicidade não são adotados, explicitamente, fatores de escala, (por convenção, parâmetro  $a=1$ ). Outra característica do modelo a fim de aumentar sua identificabilidade e interpretabilidade é imposto a condição que o somatório das

dificuldades associadas à uma mesma faceta seja igual a zero para todas as facetas, com exceção da faceta associada ao item.

#### Circunstâncias que envolvem o item

A cada circunstância que envolve o item, bem como ao próprio item pode se associar uma faceta.

- Bloco: conjunto de itens sempre aplicados simultaneamente, neste presente trabalho foram utilizados blocos contendo 11 itens
- Caderno: conjunto de blocos que compõe um teste, no presente estudo são utilizados dois blocos por caderno
- Posição do bloco: posição na qual ele é exposto no caderno (posições de 1 a 4),
- Posição da disciplina Língua Portuguesa no caderno: início ou final do caderno.

#### Efeitos possíveis de modelar

Primeiramente, deve-se conjugar as informações do entorno do problema: tem-se uma mesma população: logo qualquer subpopulação a qual seja amostra aleatória desta deve possuir a mesma média de proficiência. Nessas subpopulações, aplicam-se testes os quais inicialmente avaliam Língua Portuguesa ou inicialmente avaliam Matemática, alternando-se dentro dos cadernos a ordem dos blocos. Afinal, quais as consequências que tais práticas podem acarretar?

Visando responder a essas questões, foram propostos oito modelos de estudos, variando de duas facetas a 6 facetas conforme apresentado no Quadro II.

**Quadro II:** Modelos, facetas utilizadas e resultados encontrados nos estudos

Estudo	Proficiências	Software	Modelo	Amplitude	Facetas					
					Aluno	Item	Caderno	Posição Disciplina	Bloco	Posição do Bloco
Inicial	PL3	MIRT	3PL	0,285	X	X				
1	F_A_I	FACETS	MFRM	0,321	X	X				
2	F_A_I_B			0,320	X	X			X	
3	F_A_I_C_B_PB			0,169	X	X	X		X	X
4	F_A_I_C_B			0,166	X	X	X		X	
5	F_A_I_C			0,158	X	X	X			
6	F_A_I_C_PD			0,105	X	X	X	X		
7	F_A_I_C_PD_B			0,089	X	X	X	X	X	
8	F_A_I_C_PD_B_PB			0,088	X	X	X	X	X	X

Fonte: Elaboração própria, 2018

Nesses estudos, optou-se utilizar o programa *FACETS* (Linacre, 2013) como ferramenta de análise, pelo fato dele compreender uma ampla variedade de modelos de TRI. Especificamente, a partir do *FACETS*, pode-se trabalhar com itens dicotômicos ou politômicos em modelos multifacetados.

Tais modelos são chamados multifacetados, porque consideram não somente as características dos itens e dos respondentes, como nas análises tradicionais, mas também porque possibilitam considerar os efeitos de diversas outras facetas apresentadas no Quadro 2.

### 3. Resultados e Discussão

Para efeitos comparativos, todas as medidas de proficiência dos alunos foram padronizadas, a fim de garantir a comparabilidade entre os diversos modelos. Os estudos foram ordenados em função dos resultados das amplitudes de proficiências entre os cadernos (ordem decrescente), obtidos nos 8 modelos utilizados.

Podemos classificar, pelas amplitudes encontradas, 3 grupos com características distintas:

- O primeiro grupo caracteriza-se pelos modelos MFRM que apresentam as maiores amplitudes entre os cadernos e que se assemelham ao modelo Z3PL, são os modelos F\_A\_I e F\_A\_I\_B. Tal fato já era esperado, uma vez que o estudo 1 não considera a posição do item, e no estudo 2, como cada item pertence a um e somente um bloco – o balanceamento dos blocos pelo BIB, faz com que a inserção do bloco, como mais uma faceta no modelo, seja praticamente sem eficácia no ajuste das diferenças entre os cadernos .

- No segundo grupo encontram-se os modelos F\_A\_I\_C\_B\_PB, F\_A\_I\_C\_B, e F\_A\_I\_C. Nesses modelos, a dificuldade do caderno é a faceta relevante para um maior equilíbrio entre as proficiências, sendo praticamente irrelevante considerar ou não o bloco, pelo motivo anteriormente citado.

- No terceiro grupo, na colocação de melhor eficiência, encontram-se os modelos F\_A\_I\_C\_PD, F\_A\_I\_C\_PD\_B e F\_A\_I\_C\_PD\_B\_PB, nos quais, além de considerarmos os cadernos, inserimos as facetas posição da disciplina e posição do bloco.

Como resultado final tem-se o modelo saturado (F\_A\_I\_C\_PD\_B\_PB) como o mais eficiente por natureza, pois engloba os efeitos de todas as facetas, mas é muito menos parcimonioso. O modelo mais eficaz, obtido no estudo 6 (F\_A\_I\_C\_PD), apresenta resultados praticamente iguais aos dos estudos 7 e 8, diferenciando-se em centésimos, e demonstrando

que a posição do caderno e a ordem da disciplina no caderno são as facetas mais relevantes, uma vez que, quando essas facetas são consideradas no modelo, consegue-se atenuar as diferenças entre os diferentes cadernos de testes.

#### 4. Conclusões

Este artigo pretendeu demonstrar a versatilidade do modelo de facetas, um modelo da classe de Rasch para modelar um problema complexo com grande facilidade, o que seria muito mais difícil em um modelo de 3 parâmetros dado sua estrutura complexa por natureza. O modelo proposto torna-se útil em situações onde o foco da avaliação é o aluno em *designs* de testes compostos por diferentes modelos de cadernos e mais de uma disciplina. Procuramos demonstrar que, através da inserção de facetas apropriadas no modelo da TRI é possível estimar medidas mais próximas do desempenho real dos alunos, minimizando o efeito cansaço provocado pela ordem da disciplina nos cadernos de testes assim como a própria dificuldade inerente a cada caderno de teste.

#### Referências

- BERGAN, Jon Richard. Rach Versus Birnbaum: New Arguments in an Old Debate. Arizona, 2013.
- ECKES, Thomas. Introduction to Many-facet Rasch Measurement: Analyzing an Evaluating Rater-Mediated Assessment. Frankfurt: Peter Lang, 2011.
- ECKES, Thomas. Examining Rater Effects in TestDAf Writing and Speaking Performance Assessments: A Many – Facet Rasch Analysis. Language Assessment Quarterly, 2(3). Lawrence Erlbaum Associates, Inc. 2005.
- LINACRE, J.M. (2013). FACETS 3.71.4 [computer program anual].www.winsteps.com
- CHALMERS, R. P. A Multidimensional Item Response Theory Package for the R Environment. In: *Journal of Statistical Software*, v. 48, issue 6, may 2012.
- TOFFOLI, Sonia F. L., ANDRADE, Francisco Dalton de, BORNIA, Antonio C. Evaluation of open itens using many-facet Rasch model. Journal of Applied Statistics, 2015, DOI: 10.1080/02664763.2015.1049938.
- Wright, B.D. IRT in the 1990s: Which Models Work Best? 3PL or Rasch? Measurement Transactions, 1992.