

## **AVALIAÇÃO DA APRENDIZAGEM E DAS PROVAS DO CENTRO DE FORMAÇÃO INTERDISCIPLINAR/UFOPA VIA TEORIA DA RESPOSTA AO ITEM.**

Edilan de Sant'Ana Quaresma, ICED/UFOPA - USP(equaresm@ufpa.br)

Carlos Tadeu dos Santos Dias, ESALQ/USP(ctsdias@usp.br)

Simone Daniela Sartorio, UNIFEI (sisartorio@yahoo.com.br)

**RESUMO:** A Universidade Federal do Oeste do Pará (UFOPA), criada pela Lei nº 12.085, de 5 de novembro de 2009 apresenta na sua proposta acadêmica o Centro de Formação Interdisciplinar (CFI) com o objetivo de garantir uma “visão integrada da realidade e processos que ocorrem na natureza e sociedade – especialmente a amazônica – e que se voltam para as mais diversas expressões do desenvolvimento nas comunidades e para a promoção humana, além de prover o aluno de informações básicas sobre as ciências e institutos da Universidade”. O primeiro semestre letivo do CFI corresponde à Formação Interdisciplinar 1, universal para todos os alunos ingressantes na UFOPA e prevê a formação discente em seis ciclos, composto por seis módulos de ensino (Origem e evolução do conhecimento – OEC; Sociedade, natureza e desenvolvimento – SND; Lógica, linguagens e comunicação – LLC; Estudos integrativos da Amazônia – EIA; Seminário integradores - SINT e Interação na base real - IBR), aqui chamados de itens. A portaria da UFOPA que normatiza o progresso e percurso acadêmico define avaliação da aprendizagem como “o processo de apreciação e julgamento do rendimento acadêmico dos discentes, com o objetivo de acompanhar, diagnosticar e melhorar o processo de ensino e aprendizagem, bem como a habilitação do discente em cada componente curricular” e utiliza um Índice de Desempenho Acadêmico (IDA) baseado na média ponderada das médias finais alcançadas nos componentes curriculares, como o instrumento de mensuração do aprendizado do aluno, dentro de uma abordagem da Teoria Clássica dos Testes, baseada unicamente nos escores obtidos pelos alunos nas avaliações. Entretanto, o conjunto de respostas dadas pelos alunos aos itens da avaliação 1 do CFI pode ser utilizado para mensurar

uma variável latente (variável não mensurável diretamente, como habilidades, competências, conhecimento), aqui relacionada ao objetivo do CFI. O trabalho apresentado objetivou utilizar a Teoria da Resposta ao Item (TRI) para mensurar a variável latente, objeto do CFI, com dados provenientes de 910 alunos submetidos a todas as provas relativas à avaliação interdisciplinar 1 da UFOPA, no ano de 2011. A TRI considera o desempenho obtido pelos alunos em cada item isoladamente e não na prova como um todo, aos moldes do que hoje é feito por meio do IDA na UFOPA. Foi utilizado o modelo logístico de dois parâmetros (Birnbaum, 1968), ajustado com o auxílio do pacote ltm do software R. Os resultados identificaram o item 3 (EIA) com maior parâmetro de dificuldade e discriminação e, portanto, com menor probabilidade de acerto, ao passo que o item 6 (SINT) como o menos difícil. Analogamente, foram obtidas estimativas da variável latente (objetivo do CFI) para cada um dos alunos na avaliação 1. A TRI revelou-se como uma alternativa para a análise dos dados de avaliação da UFOPA, contribuiu para contrariar o que se pensava sobre a homogeneidade nos graus de dificuldade dos itens e propõe-se como uma alternativa viável e mais justa para avaliar alunos, considerando o desempenho dos mesmos nos itens.

Palavras-chave: TRI. Avaliação. Habilidade.

### **a) Objetivo do trabalho**

O trabalho aqui apresentado teve como objetivo geral utilizar a Teoria da Resposta ao Item (TRI) para avaliar os resultados obtidos nas provas finais da avaliação 1, do Centro de Formação Interdisciplinar da UFOPA, com dados relativos ao ano de 2011.

Dentre os objetivos específicos, elenca-se: (i) estimar e interpretar o parâmetro de dificuldade dos itens, identificando aqueles mais difíceis/fáceis; (ii) estimar e interpretar o parâmetro de discriminação dos itens; (iii) estimar e interpretar o traço latente de cada um dos alunos avaliados, relacionado ao conhecimento dos mesmos, considerando os referidos itens (provas).

## **b) Descrição do problema**

A Universidade Federal do Oeste do Pará (UFOPA) foi criada pela Lei nº 12.085, de 5 de novembro de 2009. É a primeira instituição federal de ensino superior com sede no interior do estado do Pará. Está localizada no coração da Amazônia Brasileira, dentro e fora dos limites do município de Santarém, a terceira maior cidade paraense, mundialmente conhecida por suas belezas naturais, com destaque para o exótico encontro das águas dos rios Tapajós e Amazonas, que não se misturam e pelas praias de areia branca contrastando com o azul anil do Rio Tapajós.

A UFOPA surgiu da incorporação do Campus de Santarém da Universidade Federal do Pará (UFPA) e da Unidade Descentralizada Tapajós da Universidade Federal Rural da Amazônia (UFRA), as quais mantinham atividades na região Oeste paraense. A UFOPA assimilou também outras unidades da UFPA e da UFRA para a formação dos campi de Alenquer, Itaituba, Juruti, Monte Alegre, Óbidos e Oriximiná.

A estrutura acadêmica da instituição prevê a formação acadêmica em seis ciclos, iniciados a partir de um Centro de Formação Interdisciplinar (CFI), comum a todos os que ingressam na instituição, e mais cinco institutos temáticos científicos, cujas áreas de atuação estão vinculadas ao contexto amazônico (UFOPA, 2012).

O CFI consiste de seis Módulos de Ensino, cujo objetivo é formar o discente com *“um conjunto de conteúdos articulados que se inter-relacionam para que o estudante tenha, desde o início de seu percurso acadêmico, uma visão integrada da natureza e das sociedades, especialmente amazônicas, com vistas ao seu desenvolvimento em novos patamares, mais compatíveis com os tempos futuros”*. Os módulos de ensino são: Origem e Evolução do Conhecimento (OEC); Sociedade, Natureza e Desenvolvimento (SND); Lógica, Linguagens e Comunicação (LLC); Estudos Integrativos da Amazônia (EIA); Seminários Integradores (SINT); e Interação na Base Real (IBR).

A portaria da UFOPA que normatiza o progresso e percurso acadêmico define avaliação da aprendizagem como “o processo de apreciação e julgamento do rendimento acadêmico dos discentes, com o objetivo de acompanhar,

diagnosticar e melhorar o processo de ensino e aprendizagem, bem como a habilitação do discente em cada componente curricular.” e utiliza um Índice de Desempenho Acadêmico (IDA) baseado na média ponderada das médias finais alcançadas nos componentes curriculares, como o instrumento de mensuração do aprendizado do aluno, dentro de uma abordagem da Teoria Clássica dos Testes, baseada unicamente nos escores obtidos pelos alunos nas avaliações. Entretanto, o conjunto de respostas dadas pelos discentes aos itens da avaliação 1 do CFI pode ser utilizado para mensurar uma variável latente (variável não mensurável diretamente, como habilidades, competências, conhecimento), aqui relacionada ao objetivo do CFI. Como medir esse traço latente, aqui compreendido como a habilidade individual de cada aluno, desconsiderando os procedimentos avaliativos atualmente utilizados na UFOPA, baseados na teoria clássica dos testes? O método proposto pela Teoria da Resposta ao Item pode contribuir com este processo de avaliação, mostrando-se com mais vantagens que a Teoria Clássica dos Testes?

### **c) Aspectos metodológicos e desenvolvimento do tema**

A mensuração de traços latentes teve sua origem na psicometria clássica, sendo hoje utilizada como o procedimento mais amplamente disseminado, a Teoria da Resposta ao Item (TRI). Andrade *et al.* (2000) definem a TRI como um conjunto de modelos matemáticos que procuram representar a probabilidade de um discente responder certo item (pergunta, questão) como função dos parâmetros do item e da(s) habilidade(s) do respondente, de tal forma que, quanto maior a habilidade, maior a probabilidade do sujeito acertar o item.

A TRI vem sendo utilizada com sucesso no sistema de avaliação educacional brasileiro, inicialmente no Sistema de Avaliação da Educação Básica desde sua implantação, em 1995, e a partir de 2009, no Exame Nacional de Ensino Médio (Enem). Trabalha, para dados dicotômicos, com modelos de um, dois e três parâmetros que consideram, respectivamente: (i) somente a dificuldade do item, (ii) a dificuldade e a discriminação e (iii) a dificuldade, a discriminação e a probabilidade de resposta correta dada por discentes de baixa habilidade, popularmente conhecido como “chute”.

Diferentemente da teoria clássica dos testes, em que o resultado do processo avaliativo considera apenas o escore obtido por cada aluno em um conjunto de questões, a TRI considera cada item ou questão de um instrumento avaliativo, mantendo seu foco no item individualmente e não na prova como um todo.

Para o trabalho aqui apresentado, foram utilizados resultados dos seis módulos de ensino, do Centro de Formação Interdisciplinar da Universidade Federal do Oeste do Pará, relativos à formação interdisciplinar 1, do ano de 2011, provenientes de 910 alunos submetidos a todas as avaliações, excluindo-se aqueles que deixaram de participar de pelo menos um dos processos avaliativos. Considerou-se cada módulo de ensino como um item, os quais foram dicotomizados da seguinte forma: zero para valores abaixo da média geral de cada um dos itens e um para valores a partir da média geral de cada um dos itens.

Considerou-se o modelo logístico de dois parâmetros (BIRNBAUM, 1968; HAMBLETON, SWAMINATHAN e ROGERS, 1991), definido como:

$$P(U_{ij} = 1 | \theta_j) = \frac{1}{1 + e^{-a_i(\theta_j - b_i)}}$$

em que:

- $U_{ij}$ : é variável dicotômica que assume os valores 1, quando o indivíduo  $j$  responde corretamente o item  $i$ , ou 0 quando o indivíduo  $j$  não responde corretamente o item  $i$ ;
- $\theta_j$  representa a habilidade (traço latente) do  $j$ -ésimo indivíduo, teoricamente assumindo valores no intervalo  $]-\infty; \infty[$ ;
- $P(U_{ij} = 1 | \theta_j)$  é a probabilidade condicional de um indivíduo  $j$  com habilidade  $\theta_j$  responder corretamente o item  $i$  e é chamada de Função de Resposta do Item - FRI;
- $b_i$  é o parâmetro de dificuldade do item  $i$ , medido na mesma escala da habilidade  $\theta_j$ ;
- $a_i$  é o parâmetro de discriminação do item  $i$ , servindo para identificar o quanto aquele item discrimina sujeitos com baixas habilidades de sujeitos com altas habilidades.

Para o procedimento iterativo de estimação dos parâmetros, utilizouse o pacote ltm do software R (RIZOPOULOS, 2006), que fornece uma estrutura flexível para analisar itens com estrutura dicotômica ou politômica, sob a abordagem de máxima verossimilhança marginal. Neste procedimento, utilizaram-se resultados dos valores de alfa de Cronbach para definir os itens a serem utilizados no modelo, considerando ideal aquele com o maior valor da referida medida.

Uma análise exploratória inicial, mostrada na tabela 1, indicou que o módulo de ensino SINT (item 6) apresentou maior proporção de sujeitos com resultados acima da média geral do item (61,43%), dentre aqueles com valores acima da média. Analogamente, o módulo IBR (item 5) apresentou maior proporção de sujeitos (48,90%) entre os itens analisados, dentre aqueles com resultados abaixo da média geral do item. Esta informação sugere, a princípio, que o item 6 é o mais fácil dentre os itens avaliados, ao passo que o item 5 apresenta maior grau de dificuldade.

Tabela 1 – Proporção de cada nível de resposta.  
Itens das Provas CFI/UFOPA: 2011

| Item   | 0 (abaixo da média) | 1 (acima da média) |
|--------|---------------------|--------------------|
| 1.LLC  | 0,4473              | 0,5527             |
| 2 OEC  | 0,4077              | 0,5923             |
| 3 EIA  | 0,4835              | 0,5165             |
| 4 SND  | 0,4484              | 0,5516             |
| 5 IBR  | 0,4890              | 0,5110             |
| 6 SINT | 0,3857              | 0,6143             |

Ainda na tabela 1 é possível observar que, de forma geral, os alunos obtiveram maiores escores acima da média geral para cada um dos itens avaliados, embora a diferença nos percentuais seja visivelmente pequena.

A correlação bisserial costuma ser utilizada para verificar se uma determinada variável apresenta correlação significativa com o escore bruto produzido pelo conjunto dos itens, isto é, mede a capacidade de discriminação do item em relação ao resultado do teste. Esse passo é crucial para a escolha de itens que de fato apresentam consistência interna e se associam bem ao escore que será produzido. Geralmente se aceita valores para a correlação bisserial superior a 0,3.

Os resultados encontrados para a correlação bisserial ( $LLC=0,4005$ ,  $OEC=0,3877$ ,  $EIA=0,4223$ ,  $SND=0,3085$ ,  $IBR=0,2190$  e  $SINT=0,3460$ ), mostram que somente o item 5 (IBR) não apresenta correlação bisserial superior a 0,3, indicando que este item apresenta problemas na consistência interna e que, portanto, é um forte candidato a ser retirado do processo que estima as habilidades individuais dos alunos.

O Coeficiente Alfa de Cronbach é definido como uma forma de estimar a confiabilidade de um questionário/prova aplicado a uma pesquisa/avaliação. Mede a correlação entre respostas em uma prova/questionário através da análise das respostas dadas pelos respondentes, apresentando uma correlação média entre as questões, o que pode ser interpretado como a consistência interna do teste. A tabela 2, mostrada a seguir, apresenta os valores de tal correlação, indicando o quanto cada item contribui para a correlação geral da prova.

Tabela 2 – Coeficiente Alfa de Cronbach.  
Itens das Provas CFI/UFOPA: 2011

| Item           | Valor  |
|----------------|--------|
| Todos os itens | 0,6143 |
| Menos item 1   | 0,5479 |
| Menos item 2   | 0,5534 |
| Menos item 3   | 0,5386 |
| Menos item 4   | 0,5853 |
| Menos item 5   | 0,6201 |
| Menos item 6   | 0,5703 |

Pela tabela 2 observa-se, por exemplo, que se todos os seis itens fossem considerados conjuntamente, a prova apresentaria uma correlação (aqui entendido como confiabilidade) igual a 0,6143. Entretanto, observa-se na mesma tabela, que a retirada do item 5 aumentaria a confiabilidade da prova para 0,6201, reforçando mais uma vez, a possibilidade de retirar o item 5 do processo de estimação dos parâmetros.

A retirada do item 5 foi confirmada quando se realizou um teste Qui-quadrado para verificar os itens que apresentavam correlação significativa, o qual indicou problemas na associação entre os itens 4 e 5 (valor de  $p=0,183$ ), sugerindo mais uma vez, a retirada do item 5 para estimar o traço latente individual de cada aluno, além dos parâmetros de cada item.

Todos os procedimentos aqui citados sugerindo a retirada do item 5 indicam problemas neste item e não necessariamente a sua ineficácia na contribuição da medida do traço latente. Trata-se de um item que precisava ser avaliado, possivelmente reformulado e, preferencialmente, pré-testado. Com a retirada do item 5, os resultados apresentados na tabela 2 passam a ser apresentados pela tabela 3.

Tabela 3 – Coeficiente Alfa de Cronbach Sem Item 5. Itens das Provas CFI/UFOPA: 2011

| Item           | Valor  |
|----------------|--------|
| Todos os itens | 0,6201 |
| Menos item 1   | 0,5762 |
| Menos item 2   | 0,5499 |
| Menos item 3   | 0,5394 |
| Menos item 4   | 0,5842 |
| Menos item 6   | 0,5795 |

Na tabela 3 é mostrado que a associação entre todos os itens da prova, quando se retira o item 5, passa de 61,43% para 62,01%. Embora este coeficiente meça o desempenho do instrumento em uma dada população evitando o agrupamento de questões aparentemente relevantes, este grau de associação pode aqui ser interpretado como a relação interdisciplinar na elaboração dos itens, de tal forma que quanto maior esta correlação, mais se pode dizer que os itens foram trabalhados de forma interdisciplinar. No caso dos itens em questão, infere-se que os itens foram trabalhados moderadamente de forma interdisciplinar.

O grande objetivo da Teoria da Resposta ao Item é estimar os parâmetros do modelo utilizado, no caso aqui apresentado, o modelo para dados dicotômicos com dois parâmetros para os itens, estimar a dificuldade e a discriminação de cada item, além da probabilidade de acerto do item considerando as habilidades dos respondentes. Estas informações apresentam-se elencadas na tabela 4.

Tabela 4 – Parâmetros dos itens e probabilidade de acerto.  
Itens das Provas CFI/UFOPA: 2011

| Item   | Dificuldade | Discriminação | $P(U_{ij} = 1   \theta_j)$ |
|--------|-------------|---------------|----------------------------|
| 6 SINT | -0,53546289 | 1,065660      | 0,6389065                  |
| 2 OEC  | -0,33953855 | 1,567576      | 0,6300083                  |
| 4 SND  | -0,24516307 | 1,018851      | 0,5621235                  |
| 1.LLC  | -0,21952856 | 1,238283      | 0,5675442                  |
| 3 EIA  | -0,05154374 | 1,838573      | 0,5236740                  |

A tabela 4 mostra que o item 3 (EIA) apresentou maior parâmetro de dificuldade e discriminação e, portanto, menor probabilidade de acerto, ao passo que o item 6 (SINT) apresentou-se como o mais fácil e, portanto, com maior probabilidade de acerto.

Na figura 1 são apresentadas as curvas característica e de informação dos itens, complementando as informações contidas na tabela 4. Na curva característica do item, quanto maior a inclinação, maior o poder de discriminação do item. Observe, por exemplo, que a linha verde, correspondendo ao item EIA, apresenta-se como a mais íngreme, aparentemente muito parecida com a linha de cor vermelha (item OEC), sendo considerados os itens com maior discriminação, respectivamente. Analogamente, a curva de cor azul marinho (item SND), apresenta-se como a menos inclinada, indicando o item com menor discriminação.

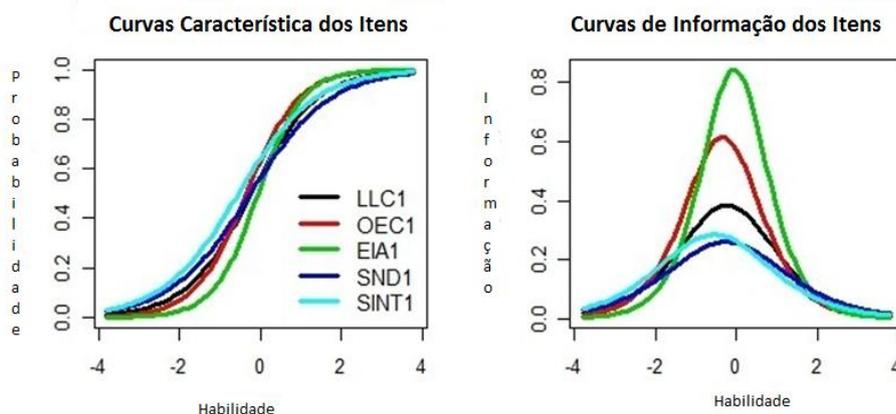


Figura 1: Curva Característica e Curva de Informação do Item

Depois de estimados os parâmetros dos itens, as habilidades dos 910 alunos foram igualmente estimadas. Tais habilidades são mostradas na tabela 5, apenas para um número reduzido de alunos, por limitação de espaço.

Tabela 5 – Estimativas das habilidades individuais de alunos.  
Provas CFI/UFOPA: 2011

| Identificação do aluno | Habilidade estimada | Erro padrão |
|------------------------|---------------------|-------------|
| xxxxxxx29              | -1,4840             | 0,6778      |
| xxxxxxx08              | -0,9236             | 0,6196      |
| xxxxxxx15              | 0,7477              | 0,6428      |
| xxxxxxx10              | 0,4987              | 0,6164      |

As estimativas das habilidades de cada um dos alunos avaliados permitem a construção de uma escala de escores e a localização de cada um dos discentes avaliados nesta escala. A grande vantagem de tal procedimento consiste na consideração do desempenho do grupo de avaliados em cada uma das cinco questões utilizadas no processo e não na média ponderada de cada prova considerada globalmente, e utilizada pela administração superior da UFOPA na classificação dos alunos.

#### **d) Conclusões**

Nas ciências sociais e humanas, a avaliação pode assumir uma grande variedade de formas, dependendo do seu objetivo. Ela pode servir como técnica de descrição de fenômenos humanos e sociais ou, assumindo aspectos mais quantitativos, servir como método de descrição entre características pessoais. A abordagem utilizada no trabalho aqui apresentado considerou alguns dos métodos estatísticos usuais nesta ampla visão de avaliação.

Tradicionalmente a avaliação do desempenho, por exemplo, de alunos, é baseada na observação da quantidade de questões respondidas corretamente dentre um conjunto total de questões, obtendo o que se conhece por escore do teste, numa abordagem conhecida como teoria clássica dos testes.

Dentre as limitações da teoria clássica dos testes, destaca-se o fato do parâmetro de dificuldade do item variar de prova para prova, dependendo da

amostra de discentes avaliados; a impossibilidade real de obter formas paralelas de um teste; a constância na variância dos erros de medidas para todos os avaliandos, dentre outras.

A TRI apresenta-se como uma alternativa interessante à teoria clássica dos testes e, neste trabalho, mostrou-se eficiente na análise de dados provenientes dos cinco módulos de avaliação do centro de formação interdisciplinar da UFOPA. Mostrou que, ao contrário do que se pensava, os módulos de ensino apresentaram-se com graus de dificuldade e discriminação diferenciados, provavelmente em função dos conteúdos avaliados em cada um dos módulos.

Foi evidenciado o caráter interdisciplinar entre os módulos de avaliação, indicando efeito positivo na proposta de formação conjunta dos discentes, voltados aos problemas da realidade local.

## Referências

- [1] ANDRADE, D.F., TAVARES, H.R., VALLE, R. da C. **Teoria da Resposta ao Item: Conceitos e Aplicações**. 14º SINAPE. 2000.
- [2] BIRNBAUM, A. **Some latent trait models and their use in inferring an examinee's ability**. In **Statistical Theories of Mental Test Scores**. (Lord, F. e Novick, M., Eds.). Reading, Mass.: Addison-Wesley. 1968.
- [3] HAMBLETON, R. K., SWAMINATHAN, H. e ROGERS, H. J. **Fundamentals of Item Response Theory**. North Caroline: Sage Publications. 1991.
- [4] RIZOPOULOS, D. **Itm: An Package for Latent Variable Modeling and Item Response Theory Analyses**. Journal of Statistical Software. v. 17, Issue 5. November, 2006.
- [5] UFOPA. **Centro de Formação Interdisciplinar**, 2012. Disponível em [http://ufopa.edu.br/index.php?option=com\\_content&view=article&id=12&Itemid=127](http://ufopa.edu.br/index.php?option=com_content&view=article&id=12&Itemid=127). Acessado em fevereiro de 2012.