

ANÁLISE DE ITENS DE UMA PROVA DE RACIOCÍNIO PROBABILÍSTICO

Elisa Henning – Marcelo Sávio Ramos – Andréa Cristina Konrath
elisa.henning@udesc.br – marcelo.ramos26@yahoo.com.br – andreak@gmail.com
Universidade do Estado de Santa Catarina, UDESC, Brasil
Universidade Federal de Santa Catarina, UFSC, Brasil

Tema: I.5 - Pensamento relacionado à Probabilidade e à Estatística.

Modalidade: CB

Nível educativo: Terciário – Universitário.

Palavras-chave: Teoria de resposta ao item. Educação Estatística. Probabilidades.

Resumo

Este artigo tem como objetivo analisar os itens e resultados de uma prova da disciplina de Probabilidade e Estatística considerando-se aspectos quantitativos e alguns qualitativos. Os aspectos quantitativos foram analisados, considerando-se propriedades estatísticas, as quais incluíram procedimentos da análise da teoria Clássica dos Testes e da Teoria de Resposta ao Item (TRI). A análise pedagógica de alguns itens compõe os aspectos qualitativos. A prova abordava o conteúdo de Probabilidade e foi aplicada a 117 estudantes de cursos de graduação de uma instituição de ensino superior. Os itens foram considerados dicotômicos, ou seja, com resposta certa ou errada. Na TRI foi aplicado o modelo logístico de dois parâmetros. Os resultados apontam que as duas abordagens concordaram quanto à dificuldade e discriminação dos itens. A questão apontada como mais difícil envolvia os conteúdos de probabilidade condicionada e eventos mutuamente exclusivos. O item mais fácil apresentava um problema de confiabilidade envolvendo eventos independentes. O item com maior poder de discriminação era relacionado ao Teorema de Bayes. Conclui-se que a TRI é uma técnica importante para a construção de instrumentos que permitam uma avaliação mais abrangente.

Introdução

A avaliação pode ser contemplada de várias formas e por meio de métodos distintos, e suas abordagens se conectam aos diversos paradigmas que se contextualizam através do tempo. Para (Rodrigues, 2006), independentemente da postura filosófica, há um eixo comum entre as diversas concepções de avaliação educacional: todas elas associam-se a um processo interpretativo, supondo um juízo de valor, qualidade ou mérito que tem por alvo diagnosticar e verificar a abrangência dos objetivos propostos no processo ensino-aprendizagem. Uma simples verificação da aprendizagem não permite concluir se tais objetivos foram alcançados, sendo que este diagnóstico vai muito além, pois há toda uma conjuntura que propicia a aprendizagem do aluno. Portanto uma avaliação mais completa deve estar embasada em um modelo que contemple todas as relações possíveis de serem avaliadas (Rodrigues, 2006).

Todavía as provas fazem parte deste processo de avaliação. E, neste sentido consideramos essencial investigar que subsídios este instrumento (a prova) pode trazer no intuito de melhorar o processo de ensino aprendizagem. Afinal, de acordo com (Borgatto e Andrade, 2012) a avaliação do desempenho dos estudantes depende fundamentalmente da qualidade dos itens da prova. Assim sendo, a Teoria Clássica dos Testes (TCT) e a Teoria da Resposta ao Item (TRI) constituem métodos estatísticos que podem contribuir para esta finalidade.

A Teoria da Resposta ao Item (TRI) supõe que o desempenho de um teste pode ser explicado por características ou variáveis latentes subjacentes do respondente, proficiências ou habilidades (Francisco, 2005). Deste modo esta metodologia permite uma melhor validação de análises das respostas, no qual parte-se do pressuposto que os itens constitutivos do instrumento de avaliação adotam a característica de “elementos centrais da análise”, em contraponto ao paradigma observado na Teoria Clássica, em que a centralização faz referência à prova como um todo (Valle, 1999).

Sob esta ótica, o objetivo principal deste trabalho é investigar aspectos referentes ao grau de dificuldade e poder de discriminação das questões de uma prova de Probabilidade e Estatística aplicada a alunos de graduação. O trabalho de Vendramini, (Silva & Canale, 2004), que aplicaram a TCT e TRI para avaliar uma prova de raciocínio estatístico motivou a realização desta pesquisa. Face à importância da Estatística como Ciência na graduação, acredita-se ser importante que o aluno compreenda os principais conceitos de probabilidade, que formam a base dos problemas de inferência. Sentimos que os alunos têm dificuldades com probabilidades, principalmente em relação aos conceitos. E, a partir dos resultados podem ser delineadas ações com objetivo de melhorar o processo de ensino aprendizagem.

Teoria Clássica dos Testes

A Teoria Clássica dos Testes (TCT) procura conhecer o comportamento que os respondentes de um teste apresentam, por meio de seu resultado final, ou seja, de seu escore total (Maia, 2009). A TCT baseia-se em parâmetros descritivos, que auxiliam na interpretação da distribuição das respostas de acordo com cada alternativa. As propriedades psicométricas dos itens de uma prova relacionam-se aos parâmetros a seguir: índice de dificuldade, índice de discriminação e correlação bisserial (Borgatto & Andrade, 2012).

O índice de dificuldade analisa o grau de dificuldade de cada item por meio da porcentagem de acerto. Quanto menor for a porcentagem de acertos, maior será o grau de dificuldade. O índice de discriminação mede a capacidade de um item de distinguir os participantes com maior habilidade (27% dos respondentes com notas mais altas) daqueles com menor habilidade (27% dos respondentes com notas mais baixas). Corresponde à diferença entre a proporção de acertos do primeiro grupo (notas mais altas) e a do segundo grupo (notas mais baixas). Já, o coeficiente de correlação bisserial é uma medida de associação entre o desempenho no item e o desempenho na prova (Borgatto & Andrade, 2012).

Teoria de Resposta ao Item

A Teoria da Resposta ao Item (TRI) é um conjunto de modelos matemáticos que procuram representar a probabilidade de um indivíduo dar uma certa resposta como função dos parâmetros do item e da habilidade do respondente. Esta função especifica que, quanto maior a habilidade, a probabilidade de uma resposta correta no ítem também é maior (Andrade, Tavares & Valle, 2000)

A idéia básica da TRI apóia-se em dois postulados fundamentais. O primeiro é que o desempenho do examinando em um item pode ser predito a partir de um conjunto de fatores (traços latentes). Outro postulado é de que a relação entre o desempenho e os traços latentes pode ser descrita por uma função monotônica crescente, cujo gráfico é chamado de Curva Característica do Item (Pasquali, 2007).

Segundo (Josemberg, 2010) os vários modelos de TRI dependem fundamentalmente do número de atributos ou dimensões assumidas (uma ou mais); do formato dos itens (por exemplo, múltipla escolha/resposta aberta, dicotômico/politômico) e do número de parâmetros dos itens a serem estimados. Na educação prevalecem os modelos logísticos de um, dois e três parâmetros para itens dicotômicos, que consideram respectivamente: somente a dificuldade do item; a dificuldade e a discriminação; a dificuldade, a discriminação e a probabilidade de resposta correta dada por indivíduos de baixa habilidade (Andrade et al. 2000).

Todavia, antes de empregar os modelos matemáticos da TRI, deve-se comprovar o cumprimento de dois pressupostos teóricos fundamentais para a utilização de modelos unidimensionais que são: o critério da unidimensionalidade e o critério da independência. Para a unidimensionalidade verifica-se se o teste possui somente um

único construto (traço latente). Já no critério da independência local, as respostas dos participantes aos itens não devem ser influenciadas pelas respostas fornecidas a outros itens, para indivíduos com uma mesma habilidade.

Como neste trabalho é usado o modelo logístico de dois parâmetros considerando itens dicotomizados, apresentamos apenas este modelo, que é sumarizado pela Equação (1):

$$P(X_{ij} = 1 | \theta_j) = P_i(\theta) = \frac{1}{1 + e^{-Da_i(\theta - b_i)}}, \quad (1)$$

onde b_i é o índice de dificuldade (ou parâmetro de posição) do item i , medido na mesma escala da habilidade θ ; a_i corresponde ao parâmetro de discriminação do item (inclinação), também chamado de parâmetro de discriminação; D denota um fator introduzido para tornar a função logística mais próxima possível da função da ogiva normal (em geral igual a 1 ou a 1,7) e n é o número de itens no teste.

PROCEDIMENTOS METODOLÓGICOS

O enquadramento metodológico utilizado deste trabalho é o mesmo de (Vendramini et al, 2004) e (Bahkta et al. 2005). Nesta pesquisa foram utilizados os dados referentes às avaliações de 117 alunos de cursos de graduação. A mesma prova foi aplicada em três turmas de Probabilidade e Estatística de três cursos diferentes: licenciatura, engenharia e computação. A prova foi composta por 5 (cinco) questões, chamadas a partir daqui de “itens”, abrangendo os seguintes assuntos, conforme Tabela 1.

Tabela 1: Assunto por item

Itens	Assunto
1	Espaço Amostral
2	Eventos mutuamente exclusivos e probabilidade condicionada
3	Regra da adição, eventos complementares e principio multiplicativo
4	Teoria de Bayes
5	Confiabilidade e eventos independentes

Fonte: Dos autores

Tanto para a análise da TCT quanto para a TRI foram consideradas as respostas dos alunos como sendo, 1 (um) o acerto, e 0 (zero) o erro. Para as análises foi utilizado o R (R Core Team, 2013), com auxílio dos pacotes ltm (Latent Trait Model) (Rizopoulos, 2006), e o psych (Revelle, 2013).

Resultados

Inicialmente foi calculada a proporção de acertos, que corresponde à proporção de participantes que responderam ao item corretamente; a proporção de erros e a correlação bisserial. Os resultados da análise clássica podem ser visualizados na Tabela 2 e os escores totais na Tabela 3.

Tabela 2: Proporção de acertos, erros e correlação bisserial

Número do item	Proporção de acertos	Proporção de erros	Correlação bisserial
1	0.5385	0.4615	0.6467
2	0.2393	0.7607	0.6141
3	0.3248	0.6752	0.6699
4	0.4957	0.5043	0.7137
5	0.6581	0.3419	0.6151

Fonte: Dos autores

Tabela 3: Escores totais

Escores Totais	0	1	2	3	4	5
Frequência	20	22	24	17	27	7

Fonte: Dos autores

O item identificado como mais difícil foi o item 2, com cerca 24% de acertos, e o mais fácil foi o item 5, com aproximadamente 66% de acertos. Para as cinco questões o escore mínimo foi zero, com vinte alunos não acertando nenhuma questão, e o máximo foi cinco, com sete alunos acertando todas as questões. Na TCT o item que aparece como o mais discriminante é o item 4. Este item apresenta a maior diferença entre a proporção de acertos dos 27% de alunos com notas mais altas e a proporção de acertos dos 27% de alunos com notas mais baixas, com valor 0,2478.

Em sequência foi feita a análise da unidimensionalidade aplicando a Análise Fatorial e Análise Paralela Modificada. Constatou-se que o teste tem um único construto. A confiabilidade da escala medida pelo coeficiente Kuder-Richardson foi de 0,672, valor pode ser considerado como aceitável, pois se aproxima da referência (0,7), sabendo-se que os itens não foram pré-testados.

Na TRI foi ajustado um modelo logístico de dois parâmetros, e as estimativas dos parâmetros de discriminação e de dificuldade podem ser visualizados na Tabela 4. Observamos que o item apresentado como o mais difícil, pela TCT, o item 2, é o mesmo pela TRI (maior valor do parâmetro de dificuldade), fenômeno que se repete para o item mais fácil (item 5).

Tabela 4: Parâmetros do modelo

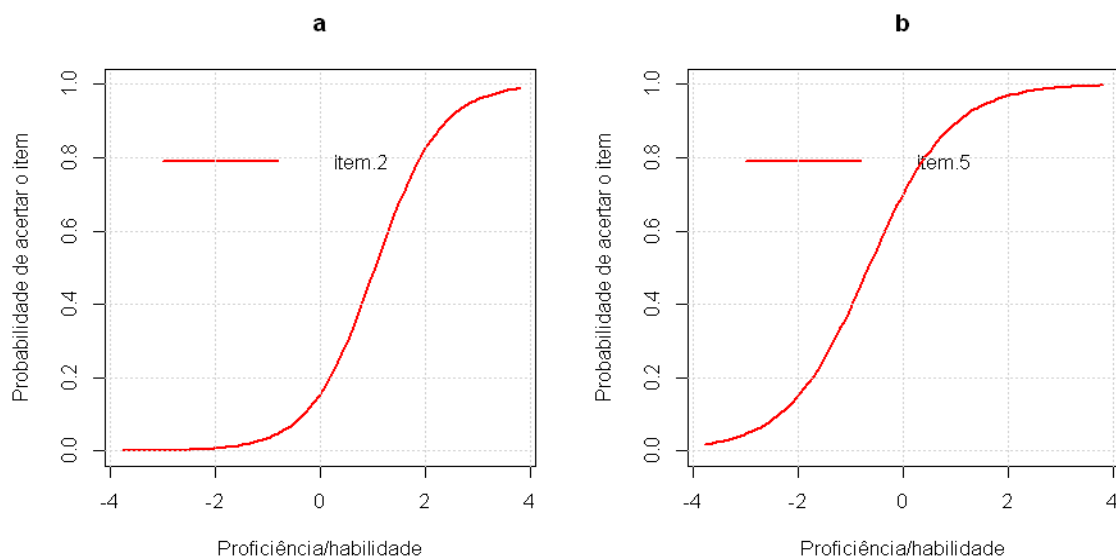
Item	Parâmetro de discriminação	Parâmetro de dificuldade
1	1.2830	-0.1515
2	1.6109	1.0407
3	1.8971	0.6253
4	2.1253	0.0269
5	1.2910	-0.6645

Fonte: Dos autores

Graus de dificuldade com valores negativos apontam para itens muito fáceis. Já para os valores positivos; valores entre zero (0) a um (1) correspondem a itens fáceis; de um (1) a dois (2) itens pouco difíceis e, a partir de dois (2) assinalam itens difíceis. Podemos considerar que o teste realizado possui itens de dificuldade moderada a muito fácil, pois os itens 1 e 5 possuem valores negativos e os demais positivos, entre a faixa de 0 a 1.

Na Figura 1 apresentam-se as curvas características dos itens 2 e 5. O item 2 (o mais difícil), Figura 1a, é relativo a eventos mutuamente exclusivos e probabilidade condicionada. Esta questão necessitava que o aluno tivesse assimilado os conceitos, o que por vezes é difícil para este.

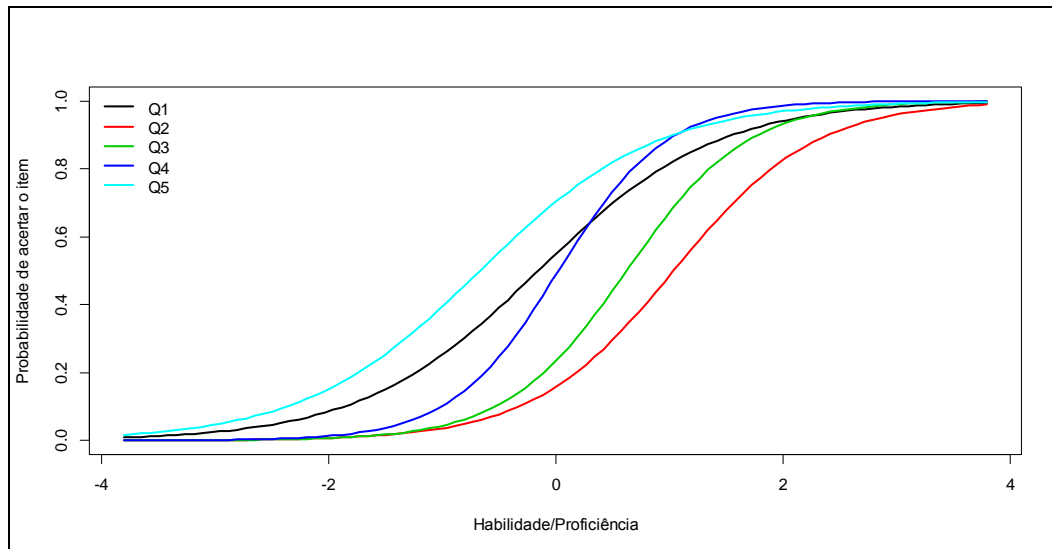
O item 5 (Figura 1b), considerado o mais fácil, corresponde à problemas de confiabilidade e eventos independentes. Na TCT apresentou percentual de 65,8% de acertos, sendo também o item mais fácil. Este classe de problemas, de confiabilidade, é fácil de relacionar com o conteúdo de Eletricidade, visto em Física Geral.

Figura 1: Curva Característica dos Itens 2 e 5


Fonte: Dos autores

A Figura 2 apresenta a curva característica dos cinco itens. De acordo com Baker (2001), os itens possuem poder de discriminação moderados a altos.

Figura 2: Curva Característica dos itens



Fonte: Dos autores

O item mais discriminante é o item quatro (4), e abrangia o Teorema de Bayes, que na nossa concepção, implica que o aluno compreenda probabilidade condicionada. A partir da Figura 2 percebe-se que a curva do item é bastante íngreme, o que significa que um deslocamento no eixo das habilidades muda significativamente a probabilidade de acertar o item. Na análise clássica, este item (4) aparece como o mais discriminante (0,2478) e apresenta também o maior valor para o coeficiente de correlação bisserial, (0,7137) como se pode visualizar na Tabela 2. Quanto ao poder de discriminação do item 5 (o mais difícil), pode-se dizer que discrimina moderadamente (Baker, 2001) com uma curva mais suave (Figura 2).

Conclusão e considerações finais

Este trabalho teve como objetivo a análise de itens e resultados à avaliação de uma prova de raciocínio probabilístico. Para atingir este objetivo foram utilizados os métodos estatísticos designados como Teoria Clássica dos Testes (TCT) e a Teoria da Resposta ao Item. Dentre diversos modelos existentes na TRI, neste trabalho foi aplicado o modelo logístico de dois parâmetros. As duas teorias apontam para as mesmas conclusões. Concluiu-se que o instrumento de avaliação utilizado possui itens pouco difíceis, mas com bom poder de discriminação.

Referências bibliográficas

- Bahkta, B., Tennant, A., Horton, M., Lawton, G. & Andrich, D. (2005). Using item response theory to explore the psychometric properties of extended matching questions examination in undergraduate medical education. *BMC Medical Education*, 5, 1-13.
- Baker, F.(2001). *The basics of item response theory* (2^a ed). Washington: Eric Clearinghouse on Assessment and Evaluation.
- Borgatto, A. & Andrade, D. (2012) Análise Clássica de Testes com diferentes graus de dificuldade. *Estudos em Avaliação Educacional*, São Paulo, 23 (52), 146-156.
- Francisco, R. (2005). *Aplicação da Teoria da Resposta ao Item (TRI) ao exame nacional de cursos (ENC) da UNICENTRO*. Dissertação de Mestrado. Programa de Pós-graduação em Métodos Numéricos para Engenharia. Universidade Federal do Paraná, Curitiba, Brasil.
- Josemberg, M.(2010). *O uso da teoria de resposta ao item em avaliações educacionais: diretrizes para pesquisadores*, 9 (3), 421-435.
- Maia, J. (2009). *Uso da Teoria Clássica dos Testes e da Teoria de Resposta ao Item na Avaliação da Qualidade Métrica de Testes de Seleção*. Programa de Pós-graduação em Educação Brasileira. Universidade Federal do Ceará, Fortaleza, Brasil.
- Pasquali, L. (2007). *Teoria de resposta ao item: teoria, procedimentos e aplicações*. Brasília: LabPAM/UnB.
- Revelle, W. (2013) *psych: Procedures for Personality and Psychological Research*, Northwestern University, Evanston, Illinois, USA, <http://CRAN.R-project.org/package=psych> . Acesso em: 23/05/2013.
- Rizopoulos, D. (2006). Ltm: Na R package for Latent Variable Modelling and Item Response Theory Analyses. *Journal of Statistical Software*, 17 (5), 1-25.
- Rodrigues, M.(2006). Proposta de Análise de itens das Provas do Saeb sob a Perspectiva Pedagógica e a Psicométrica. *Estudos em Avaliação Educacional*, 34, 43-78.
- Valle, R. (2000) Teoria da Resposta ao Item. *Estudos em Avaliação Educacional*, 21,7-88.
- Vendramini, C.; Silva, M.& Canale, M. (2004). Análise de itens de uma prova de raciocínio estatístico. *Psicologia em Estudo*, 9, 487-498.