

Análise métrica de questões componentes de testes de rendimento: mecanismo de feedback para aprimorar sua elaboração*

Wagner Bandeira Andriola**, Raimundo Helio Leite***, José Leudo Maia****

Resumo

O texto aborda a relevância da atividade de análise métrica das questões ou dos itens componentes de testes de rendimento, enquanto procedimento essencial ao *feedback* que se deve proporcionar aos elaboradores das mesmas. Nesse âmbito, o objetivo do ensaio foi comparar os efeitos das características métricas dos itens sobre a estimativa do nível de habilidade do aluno (θ), utilizando-se, para tal, o modelo de três parâmetros logísticos da Teoria da Resposta ao Item (TRI). Este trabalho mostra uma das inúmeras possibilidades resultantes do uso da TRI: análises estatísticas pertinentes para aferir se o instrumento e os itens ou questões que o compõem cumpriram suas funções pedagógicas, isto é, *se estimaram de modo válido e confiável o aprendizado dos discentes*.

Palavras chave: avaliação educacional; avaliação da aprendizagem; Teoria da Resposta ao Item (TRI); testes de rendimento.

Metric analysis of assessment test components: feedback mechanisms to improve design

Abstract

This paper overviews the relevance of item metric analysis of assessment tests as an essential feedback procedure which should be provided to those who design them. The essay intends to compare the effects of the item metric characteristics over the estimated skill level of students (θ), using the three logistic parameters of the Item Response Theory (IRT). This work shows one of the many possibilities resulting from the use of the IRT: relevant statistical analysis to verify whether or not the tool and its items meet the expected pedagogical functions, i.e.: if they can validly and reliably estimate the learning of students.

Key words: Educational Assessment; Learning Assessment; Item Response Theory (IRT); Performance Tests.

* Trabalho apresentado na III Reunião da Associação Brasileira de Avaliação Educacional (ABAVE), realizada em Belo Horizonte (maio de 2007).

** Doutor em Filosofia e Ciências da Educação (Universidad Complutense de Madrid). Atualmente é Professor Associado da Universidade Federal do Ceará (UFC); Professor do Programa de Mestrado e Doutorado em Educação; Coordenador do Mestrado Profissional em Políticas Públicas e Gestão da Educação Superior (POLEDUC/UFC); Coordenador de Avaliação Institucional (SDI/UFC); Bolsista de Produtividade em Pesquisa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). E-mail: w_andriola@ufc.br.

*** Doutor em Educação Brasileira pela Universidade Federal do Ceará (UFC), Docente Livre do Departamento de Direito Processual da Universidade Federal do Ceará (UFC). E-mail: rhleite@terra.com.br

**** Doutor em Educação Brasileira pela Universidade Federal do Ceará (UFC), Docente da Universidade estadual do Ceará (UECE).

Análisis métrico de cuestiones de tests de aprovechamiento: mecanismo de retroalimentación para perfeccionar su elaboración

Resumen

El texto aborda la relevancia de la actividad del análisis métrico de los ítems de tests de aprovechamiento, como procedimiento esencial a la retroalimentación que se debe proporcionar a los elaboradores de las preguntas. En este contexto, el objetivo del artículo fue comparar los efectos de las características métricas de los ítems sobre la estimación del nivel de aptitud del alumno (el θ), empleando el modelo de tres parámetros logísticos de la Teoría de la Respuesta al Ítem (TRI). Este trabajo presenta una de las muchas posibilidades resultantes del empleo de la TRI: análisis estadísticos pertinentes para averiguar el cumplimiento de las funciones pedagógicas por parte del test y de los ítems, es decir, si han estimado de modo válido y fiable el aprendizaje de los alumnos.

Palabras clave: evaluación educativa; evaluación del aprendizaje; Teoría de la Respuesta al Ítem (TRI); tests de aprovechamiento.

1. Introdução

Por falta de conhecimentos adequados acerca de técnicas da avaliação educacional, através do emprego de testes e provas de rendimento, muitos professores não compreendem os cuidados que devem ser tomados *antes, durante e depois da aplicação de instrumentos de medida* dessa natureza. Esses três momentos são fundamentais para que tal instrumental cumpra sua finalidade avaliativa, de modo eficaz e válido.

O primeiro momento diz respeito à aplicação rígida de normas técnicas para elaboração de questões e itens. O segundo diz respeito à aplicação do instrumento, quando se deve cuidar para que todos os alunos tenham chances iguais de responder às questões do instrumento. A terceira fase é a menos explorada pelos muitos professores. Com efeito, terminada a aplicação de um prova ou de um teste de rendimento, o professor pensa que a única coisa a fazer é preencher o mapa de notas e entregar os resultados no órgão competente. Ledo engano. Por quê? Se bem analisadas, as questões componentes das provas aplicadas podem passar a constituir um *banco de itens* e poderão vir a ser utilizadas posteriormente, conforme assevera Andriola (1998).

Nesse âmbito, as análises posteriores podem e devem se transformar numa avaliação formativa para o aluno e para o professor. Para atingir tal objetivo, é necessário que o professor empregue teorias adequadas de avaliação dos resultados de suas provas. Conforme menciona Andriola (2002), são empregadas, atualmente, duas teorias, com esse fim: A Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI).

Este ensaio mostra como o professor pode aprimorar suas provas com análises posteriores, empregando uma técnica de estimação da habilidade do aluno a partir de um modelo derivado da Teoria de Resposta ao Item (TRI) que estima três parâmetros dos itens ou das questões: a discriminação, a dificuldade e o acerto ao acaso. Ademais, esclarece a necessidade de se formar uma equipe de avaliação formada por técnicos em medidas e avaliação educacional, assessorados por especialistas no conteúdo avaliado pela prova ou teste de rendimento.

2. O que é a Teoria da Resposta ao Item (TRI)

Desde que o homem começou a tentar responder a questões sobre a natureza, de forma racional, surgiu uma linha paradigmática de investigação, tentando reduzir a uma estrutura mais unívoca à realidade observada. Os primeiros esforços nessa direção foram feitos pelos filósofos pré-socráticos. O primeiro a formular uma redução foi Tales de Mileto (624/625 a.C. - 556/558 a.C.) que entendia ser a água o elemento fundamental da natureza. Alguns anos mais tarde, Heráclito optou pelo fogo.

A propósito de Heráclito (540 a.C - 470 a.C), há uma frase desse filósofo que caracteriza muito bem a noção que ele tinha da natureza, ao expressar que ela nunca se mostra facilmente, mas, pelo contrário, a “[...] constituição das coisas gosta de ocultar-se [...]”. Depreende-se, pois, que esses filósofos já tinham idéia de que existe algo latente na natureza que é intuído pelos sentidos e pela razão. Não apenas os filósofos antigos, mas pesquisadores contemporâneos têm sido postos face a face com essa realidade.

No começo do século XIX, foi constatada que a radiação que se pode observar nos astros é apenas uma parte pequena da radiação

eletromagnética. O físico David Bohm (1917-1994), considerado um dos grandes teóricos do século passado, entendia que a realidade que se percebe é apenas uma manifestação de uma “ordem implícita” mais geral. A importância que os cientistas da atualidade dão à descoberta de uma teoria que unifique a Teoria Geral da Relatividade e a Mecânica Quântica se insere nessa perspectiva. Vê-se, portanto, que a busca de fatores ou estruturas subjacentes aos fenômenos tem sido uma constante nessa aventura da mente humana, que é a busca do conhecimento e da verdade científica.

As investigações psicopedagógicas, que trabalham na linha de identificação de estruturas ou traços latentes, se assemelham muito à visão mostrada acima; isto é, tentar identificar o que há de subjacente a uma grande massa de dados. Como se sabe, toda atividade relacionada ao ato de aplicar testes, seja de natureza psicológica ou educativa, tem como objetivo inferir traços associados aos indivíduos, deduzidos a partir das informações obtidas nos testes aplicados. É essa inferência, feita a partir de dados que representam de forma indireta o comportamento ou a aprendizagem dos sujeitos, que se constituiu um dos grandes problemas associados com o processo de medir (ANDRIOLA, 2006).

Ao contrário das medidas físicas, obtidas através de processos diretos de medição, as educacionais ou psicológicas são conseguidas de forma indireta. Apesar dessa dificuldade, desenvolveu-se, no campo educacional, uma teoria para avaliar a qualidade dos testes de rendimento escolar, baseada em indicadores da qualidade individual dos itens e na determinação de um coeficiente avaliador do erro associado à medida do instrumento. Na área dos testes psicométricos, desenvolveu-se, nas últimas décadas, uma teoria que se denominou de Teoria do Traço Latente, Modelos do Traço Latente ou Modelos Estruturais Latentes, cuja origem está nos trabalhos pioneiros de Lawley (1943), Tucker (1946), Lord (1952, 1953), e Birnbaum, (1957, 1958).

3. Os modelos usados pela Teoria de Resposta ao Item (TRI)

Essa teoria tem pressupostos que devem ser cumpridos, como a unidimensionalidade do traço a ser medido e a invariância dos parâmetros do modelo empregado. Existem modelos matemáticos usados pela TRI que medem um parâmetro (RASCH, 1960), de dois parâmetros de por Birnbaum (1957, 1958, 1968) e de três parâmetros (LORD, 1980).

Nesse contexto, vale a pena enfatizar o principal suposto da TRI: o da unidimensionalidade. Ao referido princípio subjaz uma idéia simples e atraente, que podemos encontrar em antigas concepções gregas, conforme assevera Dunham:

A los antiguos griegos les apasionaban las simetrías, la belleza visual y la sutil estructura lógica de la geometría. Especialmente fascinante les resultaba en como lo simple y elemental podían servir de fundamento a lo complejo e intrincado.

(2002, p. 34)

A unidimensionalidade é uma proposição teórica parcimoniosa e elegante, segundo a qual toda a complexidade intrínseca ao ato de resolução de um problema —de natureza cognitiva ou não— deve ter como causa uma única estrutura latente, denominada θ (ANDRIOLA, 2006). Sendo assim, *existirá uma relação funcional entre θ e os padrões das respostas dadas a um problema*, sendo essa a idéia fundamental dos modelos logísticos de TRI, conforme destacam Hambleton (1983, 1990), Van der Linden e Hambleton (1997).

4. Objetivo do estudo

Nesse âmbito, o objetivo deste ensaio é comparar os efeitos das características métricas dos itens sobre a estimativa do nível de habilidade do aluno (θ). Para esse caso, adotamos itens¹ com índices

I Utilizam-se, neste estudo, questões da prova de Português do vestibular 98.I da Universidade Federal de Roraima, que foram analisadas por Leite (2006), em sua tese de Livre Docente. Nesse estudo, Leite (2006, p.74) determinou a unidimensionalidade de quinze, das trinta questões da prova, determinando a existência de um fator predominante.

de discriminação, de dificuldade e acerto casual, calculados pelo modelo TRI de três parâmetros, considerando três situações: 1) índices com os três parâmetros semelhantes; 2) índices com os três parâmetros diferentes; 3) índices com diferença na discriminação e na dificuldade, mas com acerto casual semelhante.

5. Apresentação e análise métrica das questões

As estimativas de habilidade serão realizadas, usando-se o *software* Gráficos *ITR.xls* de José Luiz Gaviria². Esse *software* estima a pontuação de um aluno, comparando o seu desempenho em duas questões, proporcionando, desse modo, verificar a existência de funcionamento diferencial (DIF) no item em foco (ANDRIOLA, 2002).

Trata-se, portanto, de ferramenta extremamente útil para os elaboradores de testes, que passam a dispor de informações em várias questões ao mesmo tempo. Com esse tipo de informação, o elaborador pode verificar quais as questões que poderão ser incluídas num teste, a partir do estudo comparativo dos efeitos sobre a habilidade do aluno nessas questões. Neste estudo, para fins de simplificação, serão feitas comparações apenas entre duas questões.

No cálculo da estimativa da habilidade do aluno, Gaviria Soto utiliza o modelo de três parâmetros, no qual:

- a - índice de discriminação;
- b - índice de dificuldade;
- c - chance de acerto ao acaso.

Na macro, Gaviria Soto intruduz o parâmetro u , isto é: se o sujeito contestou o item correto (acertou para 1 e errou para 0). A linha mais abaixo que aparece nos gráficos é a função de verossimilhança, que estima a habilidade do aluno. A utilidade e a inovação do trabalho de Gaviria Soto advêm do fato da rotina calcular a provável habilidade combinada de acertos e erros em mais de um item.

2 Departamento de Métodos de Investigación y Diagnóstico, Facultad de Educación, Universidad Complutense de Madrid (UCM).

No que concerne à análise dos três parâmetros das questões, calculados pela Teoria de Resposta ao Item (TRI), será feita de acordo com critérios empregados por Leite (2006, p.75), a saber:

I) A discriminação a de cada questão levará em consideração as seguintes categorias apresentadas a seguir.

Para a determinação das faixas para análise do parâmetro de discriminação, considerou-se seu maior valor, isto é, $a = 2,765$, e dividiu-se esse valor por 3. A discriminação de cada questão levará, em consideração as seguintes categorias:

| | |
|---|-------------------|
| Questões com baixo poder de discriminação | $a < 0,92$ |
| Questões com médio poder de discriminação | $0,92 < a < 1,85$ |
| Questões com alto poder de discriminação | $a > 1,85$ |

II) Para a análise do parâmetro b , adotou-se o mesmo procedimento, tendo sido observado o maior $b = 7,182$. Adotando-se o mesmo procedimento anterior, e arredondando-se para baixo as decimais do valor do intervalo, tem-se que, para a dificuldade de cada questão, os intervalos:

| | |
|-----------------------------|----------------------|
| Itens fáceis | $b < -2,394$ |
| Itens com dificuldade média | $-2,394 < b < 2,394$ |
| Itens difíceis | $b > 2,394$ |

III) o índice de acerto ao acaso c depende do número de alternativas. No caso da prova de Português em análise, as questões tinham cinco alternativas. Logo, o erro teórico aceitável é até 20% de escolhas aleatórias.

A pergunta focal que deve fazer um elaborador de teste é: quais questões devem ser escolhidas para compor um teste, tendo em vista a abrangência de conteúdo que se quer avaliar e o nível de dificuldade, discriminação e acerto casual desejados? As discussões a seguir responderão a essa inquietação enfocando, para tal, duas questões simultaneamente.

5.1. Índices com os três parâmetros semelhantes

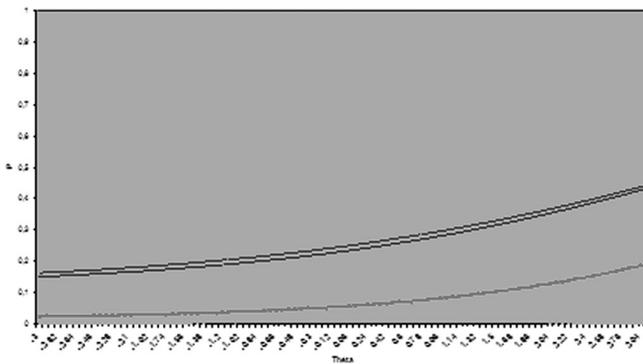
Ao se trabalhar com índices semelhantes a comparação fica mais evidente no tocante aos efeitos do erro e do acerto na estimativa da habilidade do aluno, uma vez que os parâmetros têm a mesma ordem de grandeza. Tome-se para comparação, as questões 4 e 5, cujos parâmetros são mostrados a seguir.

Quadro I. Parâmetros das questões 4 e 5 da prova de Português.

| | a | b | c |
|-----------|-------|-------|-------|
| Questão 4 | 0,413 | 4,356 | 0,124 |
| Questão 5 | 0,436 | 4,341 | 0,117 |

Essas duas questões têm baixos índices de discriminação, apresentam altos índices de dificuldade, e os índices de acerto ao acaso estão aceitáveis por se situarem abaixo do nível teórico aceitável. No entanto, qual a questão que deve ser escolhida para compor uma prova? Para responder à indagação, o gráfico I revela o comportamento da estimativa da habilidade de um aluno que responda, corretamente, às duas questões.

Gráfico 1 Estimação da pontuação Theta de um aluno que acerta as questões 4 e 5



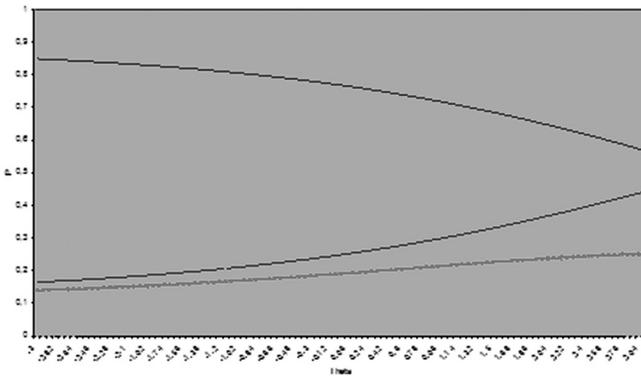
- Convenção:
- Função de verossimilhança.
 - Questão 4.
 - Questão 5.

Os gráficos confirmam o fato de os índices serem muito assemelhados, revelando que quase não há diferença de habilidade entre as duas

questões. A curva em vermelho, que representa a habilidade estimada pela função de verossimilhança, segue o mesmo padrão das curvas das duas questões, mostrando-se, todavia, num patamar de probabilidade inferior às duas questões individualmente. Por outras palavras, ele segue o mesmo padrão das curvas das questões que a geraram.

O que ocorre se o aluno acerta a questão 4 e erra a questão 5? O Gráfico 2 mostra a estimação da curva de habilidade θ , quando ao aluno acerta a questão 4 e erra a questão 5.

Gráfico 2 Estimação da pontuação Theta de um aluno que acerta a questão 4 e erra a questão 5



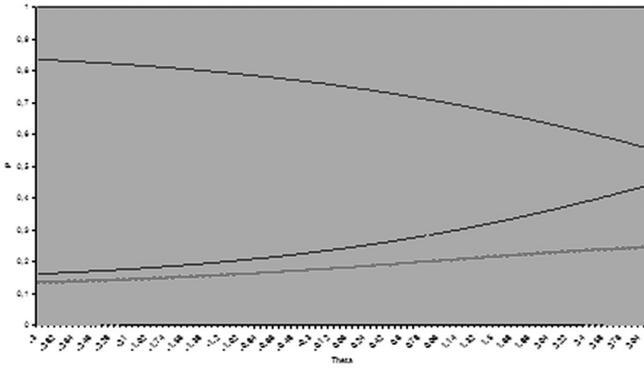
Convenção:

- Função de verossimilhança.
- Questão 4.
- Questão 5.

Observam-se dois efeitos contrários: a) a habilidade do aluno, que está alta nos níveis inferiores de θ , cai, gradativamente, nos níveis mais altos da referida variável latente; b) ocorre o efeito inverso em relação à questão 5. O resultado disso é que a habilidade estimada sofre um empuxo para cima, mas se estabiliza nos níveis mais altos de θ .

Veja-se agora o que acontece se aluno erra a questão 4 e acerta a questão 5, através do Gráfico 3.

Gráfico 3 Estimação da pontuação Theta quando o aluno erra a questão 4 e acerta a questão 5



Convenção:

- Função de verossimilhança.
- Questão 4.
- Questão 5.

O gráfico 3 tem praticamente a mesma conformação que o Gráfico 2, de modo que errar ou certar uma dessas duas questões tem efeito invariante na estimativa da habilidade do aluno. Logo, o elaborador pode escolher qualquer uma das questões. Nesse caso, o critério da abrangência de conteúdos a serem medidos deve prevalecer no processo de tomada de decisão acerca do emprego de uma ou de outra das duas questões em foco.

5.2 Índices com os três parâmetros diferentes

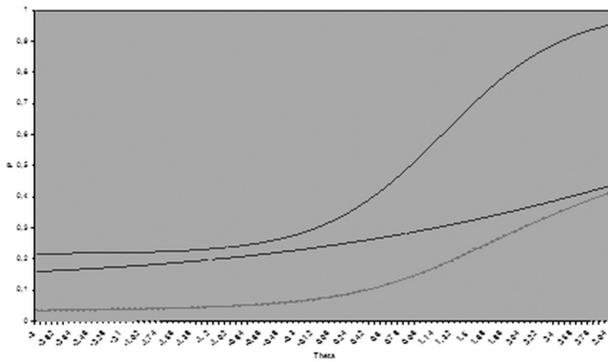
A situação mais comum é a mostrada no Quadro 2, em que os três parâmetros são diferentes.

Quadro 2. Parâmetros das questões 3 e 4 da prova de Português.

| | a | b | C |
|-----------|-------|-------|-------|
| Questão 3 | 1,588 | 1,245 | 0,216 |
| Questão 4 | 0,413 | 4,356 | 0,124 |

A questão 3 tem discriminação e dificuldade média, com acerto casual ligeiramente acima da média. A questão 4 discrimina medianamente, mas exibe alto nível de dificuldade e cerca da metade do valor teórico esperado para o acerto casual, conforme demonstrado no Gráfico 4.

Gráfico 4 Estimação da pontuação Theta de um aluno que acerta a questão 3 e acerta a questão 4



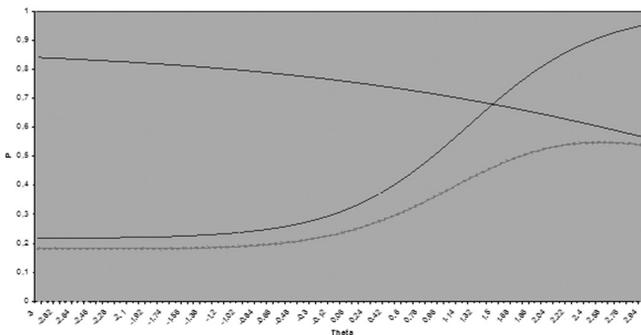
Convenção:

- Função de verossimilhança.
- Questão 3.
- Questão 4.

Nesse caso, a habilidade estimada é, predominantemente, influenciada pela questão 4, mas sofre o efeito restritivo da questão 3, o que significa que a habilidade estimada jamais ultrapassará a curva da questão 3. Entra aqui, o papel de especialista no conteúdo da prova: por que está acontecendo esse fenômeno? Qual a relação cognitiva que existe entre os conteúdos das duas questões?

O gráfico mostra o efeito do fato de o aluno errar a questão 4.

Gráfico 5 Estimação da pontuação Theta de um aluno que acerta a questão 3 e erra a questão 4



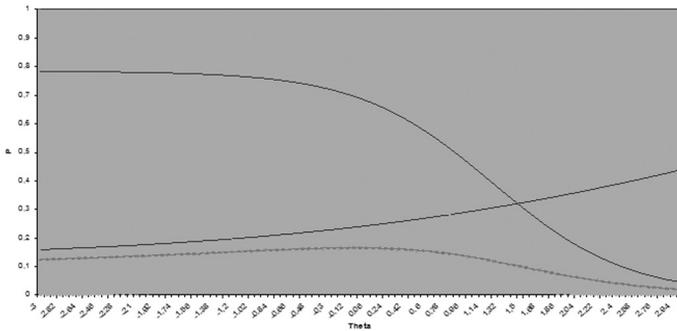
Convenção:

- Função de verossimilhança.
- Questão 3.
- Questão 4.

Nesse caso, a habilidade estimada continua seguindo o padrão de efeitos vistos antes, só que o padrão da curva estimada de habilidade muda totalmente. Ela apresenta um ponto de máxima habilidade em torno de $\theta = 2,60$ e começa, a partir desse valor, a declinar.

Quando o aluno erra a questão 3, o efeito é mais devastador na estimativa de sua habilidade, como mostra o gráfico a seguir.

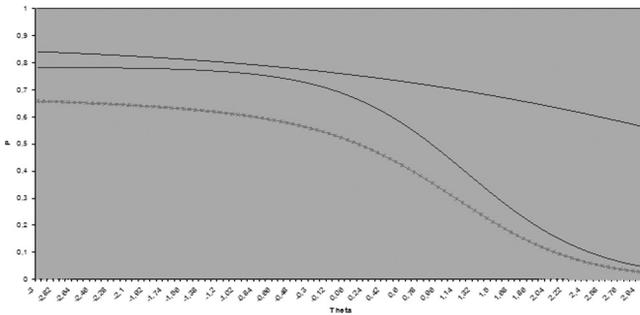
Gráfico 6 Estimação da pontuação Theta de um aluno que erra questão 3 e acerta a questão 4



- Convenção:
- Função de verossimilhança.
 - Questão 3.
 - Questão 4.

Observa-se que o nível da habilidade estimada predomina nos níveis baixos de θ . Esta análise sugere a manutenção da questão 3, em lugar da questão 4, visto que ela solicita maiores níveis de habilidade do aluno avaliado. Mostra-se a seguir, para fins didáticos, o gráfico quando o aluno erra as duas questões.

Gráfico 7 Estimação da pontuação Theta de um aluno que erra questão 3 e acerta a questão 4



Convenção:

- Função de verossimilhança.
- Questão 3.
- Questão 4.

Como era de esperar, a habilidade é mais eficazmente estimada para os níveis baixos de θ . Sendo assim, há de se indagar acerca da relevância pedagógica do emprego dessa questão em situação de avaliação do aprendizado discente.

5.3 Índices com diferença na discriminação e na dificuldade, mas com acerto casual semelhante

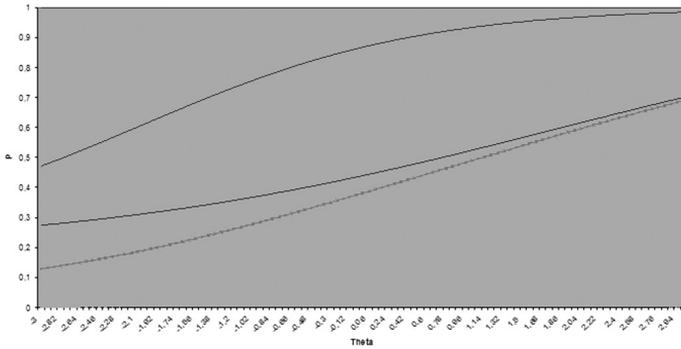
No Quadro 3 encontra-se duas questões com diferentes valores para a discriminação, para a dificuldade, porém com valor do acerto ao acaso muito semelhante.

Quadro 3. Parâmetros das questões 9 e 8 da prova de Português.

| | a | b | c |
|-----------|-------|--------|-------|
| Questão 9 | 0,773 | -2,059 | 0,214 |
| Questão 8 | 0,480 | 1,81 | 0,187 |

Observa-se que as questões 9 e 8 apresentam discriminação média, mas são diferentes no que diz respeito à dificuldade. A questão 9 é fácil, e a questão 8 possui índice médio de dificuldade. As duas exibem valores do acerto casual assemelhados.

Gráfico 8 Estimação da pontuação Theta de um aluno que acerta a questão 9 e acerta a questão 8

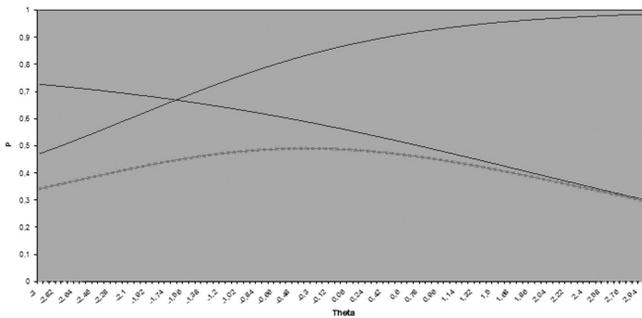


Convenção:

- Função de verossimilhança.
- Questão 9.
- Questão 8.

A habilidade estimada segue um padrão crescente, mas limitado pela curva da questão 8. O que ocorre se um aluno erra a questão 8? O gráfico a seguir retrata essa situação.

Gráfico 9 Estimação da pontuação Theta de um aluno que acerta a questão 9 e erra a questão 8

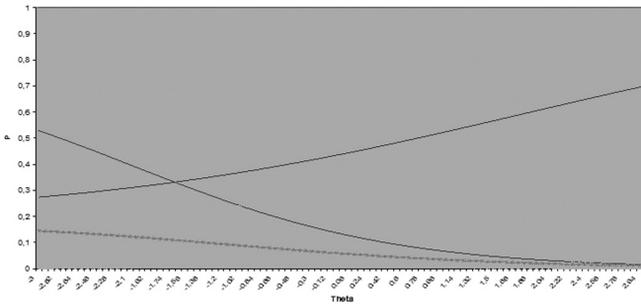


Convenção:

- Função de verossimilhança.
- Questão 9.
- Questão 8.

Acertar a questão 9 e errar a questão 8 tem efeito sobre a estimação da habilidade do indivíduo. Observa-se que a função de verossimilhança cresce até atingir um ponto máximo em torno de $p = 0,45$, e decresce nos níveis mais altos da variável latente θ . O gráfico a seguir traz o inverso dessa situação.

Gráfico 10 Estimação da pontuação Theta de um aluno que erra a questão 9 e acerta a questão 8



Convenção:

- Função de verossimilhança.
- Questão 9.
- Questão 8.

Vê-se que o nível de habilidade estimada de um aluno que erra a questão 9 e acerta a questão 8 é maior nos pontos baixos de θ , mas decresce firmemente nos níveis mais altos de habilidade. Nesse caso, a questão 9, em detrimento da 8, deve ser a escolhida para compor a prova.

Considerações finais

Como dito, há entre professores quem questione a finalidade de se fazer análises posteriores dos resultados de provas aplicadas em situações de avaliação do aprendizado discente. Espera-se que este ensaio tenha contribuído para desfazer essa falsa impressão, oriunda da falta de informações corretas sobre técnicas de avaliação de provas ou testes de rendimento.

A atividade de avaliação de um instrumento de rendimento escolar começa **antes** de sua elaboração e continua **após** seu emprego em situação de verificação do progresso discente. A fase anterior à aplicação exige que o elaborador discrimine numa tabela os conteúdos que deseja avaliar e os níveis nos quais esses conteúdos deverão ser aferidos, e validar esses conteúdos via comparação criteriosa com a abordagem com a qual foram dados em sala de aula. Após a aplicação, devem ser feitas as análises pertinentes para aferir se o instrumento e os itens ou questões que o compõem cumpriram suas funções pedagógicas: *estimar de modo válido e confiável o aprendizado dos discentes.*

Este trabalho mostra uma das possibilidades disponíveis. Evidentemente que se fizeram apenas três simulações com um pequeno número de questões. O procedimento exemplificado pode ser aplicado para se escolher quantas questões se queiram, impondo as qualidades de discriminação, dificuldade e acerto casual que o elaborador desejar.

A análise mostrou, igualmente, a necessidade da formação interdisciplinar de um grupo de avaliação, que deve ser formado, idealmente, por especialistas de conteúdo e de técnicos em medidas educacionais. Há que se desenvolver um trabalho sinérgico entre esses dois tipos de especialista para que uma equipe de avaliação desenvolva mais eficientemente seu trabalho.

Lembra-se, por oportuno, que os modelos empregados na Teoria de Resposta o Item (TRI) oferecem ampla possibilidade de se analisar também as escalas de medida do traço latente (θ).

Referências Bibliográficas

- ANDRIOLA, W. B. Utilização da Teoria da Resposta ao Item (TRI) para a organização de um banco de itens destinados à avaliação do raciocínio verbal. *Psicologia Reflexão e Crítica, Rio Grande do Sul*, v. 11, n. 2, p. 295-308, 1998.
- _____. *Detección del Funcionamiento Diferencial de los Ítems (DIF) en tests de rendimiento. Aportaciones teóricas y metodológicas (Tesis Doctoral)*. Madrid: Editora de la Universidad Complutense de Madrid, 2002.
- _____. Tendências da Psicometria Moderna. In: I Reunião da Associação Brasileira de Avaliação Educacional (ABAVE), 2006, Belo Horizonte: ABAVE, 2006.
- BIRNBAUM, A. Efficient design and use test of mental ability for various decision making problems. *Series Report n.58-I6*, Project no 7755-23, USAF School of Aviation Medicine, Randolph Air force Base, Texas, 1957.
- _____. On estimation of mental ability. *Series Report n.I5*, Project no 7755-23, USAF School of Aviation Medicine, Randolph Air force Base, Texas, 1958.

- _____. Some latent trait models and their use in inferring an examinee's ability. In: LORD, F. M. and NOVICK, M. R. *Statistical theories of mental scores (chapters 17-20)*, Reading, MASS: Addison - Wesley, 1968.
- DUNHAM, W. (2002). *Viaje a través de los génios. Biografías y teoremas de los grandes matemáticos*. Madrid: Pirâmide.
- HAMBLETON, R. K. Application of item response models to criterion-referenced assessment. *Applied Psychological Measurement*, 7 (1), 33-44, 1983.
- _____. item response theory: introduction and bibliography. *Psicothema*, II (1), 97-107, 1990.
- LAWLEY, D. N. On problems connected with item selection and test construction. *Proceedings of the Royal Society of Edinburgh*, p. 273-287, 1943.v.61.
- LEITE, R. H. *Uma aplicação da Teoria de Resposta ao Item: Processo, Resultados e Mudanças no Vestibular da Universidade Federal de Roraima*. Fortaleza, Faculdade Educação da Universidade Federal do Ceará, Tese de Docência Livre, 2006.
- LORD, F. M. A theory of test scores. *Psychometric Monograph*, n.7, Iowa City, 1952.
- _____. An application of confidence intervals of maximum likelihood to the estimation of an examinee's ability. *Psychometrika*, p.57-75, 1953.v.18.
- _____. *Application of item response theory to practical problems*. Hillsdale, N. J: LEA, 1980.
- VAN der LINDEN, W. J. & HAMBLETON, R. K. (1997). *Handbook of Modern Item Response Theory*. New York: Springer-Verlag Inc.