

O uso da informática na avaliação por métodos probabilísticos: estudo de caso em um curso de geometria analítica e vetores

Karina Pereira Carvalho¹ e Niltom Vieira Junior²

RESUMO

Este trabalho faz uso de um método probabilístico, baseado na Teoria de Resposta ao Item, para a aplicação de testes informatizados. Em campo, foram comparados os resultados de uma prova tradicional com uma avaliação informatizada em uma turma de estudantes de licenciatura em matemática. Os resultados e as dificuldades encontradas nesta prática são discutidos neste trabalho.

Palavras-chave: informática na educação, geometria analítica e vetores, testes adaptativos informatizados.

1 INTRODUÇÃO

A história da informática no ensino se confunde, mesmo que involuntariamente, com a história da informática na avaliação. Segundo Cattai (2009) o ensino através de informática tem suas raízes através das máquinas e os primeiros registros que se tem, remontam à ideia de Sidney Pressey que, em 1924, arquitetou uma máquina (mecânica) para correção de testes de múltipla escolha.

Segundo Vieira Junior (2012) um teste adaptativo informatizado é um método no qual são administrados diferentes níveis de dificuldade de questões, de acordo com a habilidade de cada indivíduo, assim podendo verificar a evolução da aprendizagem individual do aluno. Quando é usado para avaliar concede, por exemplo, um sistema de pesos (não necessariamente linearizado) que leva em consideração a habilidade do aluno. Às vezes uma questão com dificuldade menor que a habilidade do aluno vale mais do que uma questão de dificuldade muito alta (afinal ele não tem "obrigação" de acertar questões extremamente difíceis, portanto, elas têm peso menor).

Oliveira (2002) afirma que os testes adaptativos informatizados têm uma vantagem maior em relação àqueles tradicionais (lápiz e papel), pois, é possível encontrar em menos tempo, um resultado superior ou equivalente ao dos testes convencionais. Aliado a isto, aponta que tais testes tem evoluído significativamente, utilizando inclusive procedimentos probabilísticos como a Teoria de Resposta ao Item (TRI).

Através da TRI é possível verificar tais habilidades por meio da correspondência entre a pontuação obtida em um teste e os itens fornecidos a um estudante, pois, trata-se de um conjunto de modelos matemáticos usados para fazer previsões, estimativas ou inferências sobre as habilidades (ou competências) medidas de um teste. (HAMBLETON; SWAMINATHAN, 1985, apud OLIVEIRA, 2002).

¹ Licenciada em matemática pelo Instituto Federal de Minas Gerais, mestrando em matemática pela Universidade Federal de Itajubá (karinapkarvalho@hotmail.com).

² Professor de matemática no Instituto Federal de Minas Gerais, doutor em engenharia pela Universidade Estadual Paulista (niltom@gmail.com).

2 O MÉTODO ADOTADO

Este trabalho fez uso do *software* Sophia, apresentado por Vieira Junior (2012), que além de ser aplicável ao ensino informatizado possui também um módulo dedicado às avaliações.

Sua lógica interna opera com o modelo de TRI de um parâmetro, que define que a probabilidade $P_i(\theta)$ de um aluno com habilidade θ responder uma questão de dificuldade b_i corretamente, como:

$$P_i(\theta) = \frac{1}{1 + \exp^{-1(\theta - b_i)}} \quad (1)$$

Sendo a habilidade inicial medida também pelo mesmo *software* através de uma sequência iterativa de questões com diferentes níveis. Quando o estudante acerta uma pergunta, a próxima a ser administrada pelo sistema é mais difícil. O Inverso ocorre quando há um erro. Após uma série de itens tem-se estabelecido a habilidade inicial (esta verificação se dá no início do curso, para quantificar a habilidade antes e depois da etapa de estudo).

Ao fim do período de aulas, quando da aplicação da prova, de posse da equação (1) o sistema calcula também o que Santos et al. (2004) chamam de “esperança de acerto”:

$$E(x) = b_i \cdot P_i(\theta) \quad (2)$$

E então, a nota do estudante, após responder uma série de perguntas informatizadas (criadas pelo professor), é dada pela razão entre o desempenho obtido pelo desempenho esperado (conforme seu nível de habilidade):

$$Nota = \frac{\sum_{i=1}^n R_i \cdot E_i(x)}{\sum_{i=1}^n E_i(x)} = \frac{\sum_{i=1}^n R_i \cdot b_i \cdot P_i(\theta)}{\sum_{i=1}^n b_i \cdot P_i(\theta)} \quad (3)$$

Tendo que, R_i representa a resposta do item (0 ou 1, que significa errado ou certo).

Em outras palavras, este procedimento permite verificar, ao menos, duas condições que são omitidas por avaliações tradicionais:

- quando um estudante, na avaliação ao final da etapa, acerta além das questões mais fáceis do que previsto, pela sua habilidade inicial, também questões mais difíceis, tem-se um registro claro de que houve evolução conceitual no seu entendimento do conteúdo. Por esta razão, o sistema de pesos atribuído pelo método probabilístico dá a ele uma valoração maior da nota recompensando sua evolução notória;
- quando um estudante, na avaliação final, acerta questões de dificuldades aleatórias (por exemplo, errando as mais fáceis e acertando diversas questões mais difíceis, em comparação a sua habilidade inicial) o sistema interpreta um comportamento atípico, que pode indicar algum tipo de fraude. Deste modo, o sistema de pesos reduz a nota final este aluno.

3 ESTUDO DE CAMPO

Em uma turma com 14 estudantes de licenciatura em matemática do Instituto Federal de Minas Gerais - Formiga, durante a disciplina “geometria analítica e vetores” e com o aval do docente responsável, resolveu-se empregar o método probabilístico informatizado com o intuito de comparar seus resultados com aqueles obtidos pela prova convencional aplicada pelo professor.

Os conteúdos abordados pela disciplina, no intervalo de realização da pesquisa, foram:

- Semana 1: grandezas escalares e vetoriais; definição de segmentos, vetores e versores; vetores ortogonais, colineares e coplanares; adição e subtração de vetores no plano (R^2); multiplicação de vetores por escalares; ângulo entre vetores;
- Semana 2: combinação linear; base canônica; coordenadas cartesianas e polares; módulo de vetores; operações no espaço (R^3); octantes e planos;
- Semana 3: vetores definidos por dois pontos; vetor posição; representantes de vetores; paralelismo de vetores; ponto médio.

A cada semana eram realizadas 4 aulas (de 50 minutos cada) sendo duas teóricas e duas em laboratório (onde os conteúdos estudados eram vistos computacionalmente através dos *softwares* Geogebra e Winplot para reforço da aprendizagem).

O período de estudo teve duração de quatro semanas, sendo as três primeiras semanas dedicadas a apresentação de conteúdos pelo professor e a última semana a resolução de exercícios extras.

Após cada módulo de 4 aulas teóricas (cada semana) o professor entregava aos alunos uma lista de exercícios, a qual eles faziam e entregavam na aula seguinte. Ao longo do acompanhamento das aulas foram distribuídas três listas, que juntas somavam 136 itens. Ao final desta etapa foi aplicada, pelo professor, uma prova escrita (a qual seria, tradicionalmente, seu principal método avaliativo somado a um percentual mínimo de pontuação para quem lhe entregasse todas as listas de exercícios). Este processo gerou uma nota, aqui definida como “tradicional”, que foi atribuída pelo professor da turma em análise. Ressaltando que, nesta prova, foram considerados apenas erros e acertos sem nenhuma outra observação diferenciada (prova convencional).

Ao mesmo tempo e paralelamente foi realizada, com fins comparativos, a avaliação através do *software* Sophia que utiliza a TRI e os resultados são apresentados a seguir.

4 RESULTADOS

Se comparados os resultados que o professor titular da disciplina iria atribuir a mesma turma (usando apenas seus métodos habituais de correção de prova aliados a pequena pontuação pela entrega de trabalhos) com os resultados obtidos pela avaliação informatizada ter-se-ia o confronto visto na Tabela 1.

Tabela 1 – Comparação de resultados.

NOTA FINAL		
ALUNOS	TRI	Avaliação tradicional
Aluna 1	78,55*	73,35
Aluna 2	70,67*	65,02

Aluno 3	73,33*	64,52
Aluna 4	53,59	54,70
Aluna 5	68,66*	65,20
Aluno 6	46,30	49,26
Aluna 7	93,06*	88,76
Aluna 8	80,23	82,38
Aluno 9	70,34*	67,11
Aluna 10	38,52*	31,88
Aluna 11	74,50*	64,97
Aluno 12	59,24*	54,41
Aluna 13	86,33*	78,41
Aluno 14	80,67*	77,08
* quando a nota informatizada foi maior que a tradicional		

Fonte: os próprios autores.

Observou-se que em 11 casos a avaliação informatizada atribuiu maior nota aos estudantes, comparada com a avaliação que tradicionalmente seria realizada pelo professor, e em 3 casos a avaliação informatizada reduziu a nota.

Pode-se, de imediato, verificar que os casos em que a nota subiu foram atribuídos ao fato deste método considerar a “evolução” individual do estudante antes e depois dos estudos. A valorização do desempenho do aluno e o quanto ele progrediu não é feita por outros métodos. Há, porém, um problema: a possibilidade da TRI atribuir uma nota alta para um aluno, que mesmo que tenha evoluído bastante perante sua condição inicial, ainda possua conhecimentos restritos sobre o conteúdo (o que não justificaria sua aprovação).

Ao se analisar individualmente cada estudante obteve-se o seguinte:

- Aluna 1: embora displicente em sala de aula, a aluna sempre se mostra estudiosa extra-sala e o método probabilístico valorou sua significativa evolução perante a habilidade medida no início do curso, portanto, considera-se que o procedimento obteve sucesso;
- Aluna 2: em um caso similar ao anterior a aluna apresentava problemas de concentração em sala, porém, demonstrou grande evolução até a realização da prova o que fez valer a valorização dada pelo teste informatizado;
- Aluno 3: idem aos casos anteriores;
- Aluna 4: para fins de comparação realizou-se após a prova uma entrevista guiada com os alunos, com base na própria prova, com fim de verificar algum traço latente de conhecimento (para mais ou para menos em relação a nota da prova). Para esta aluna observou-se que muitos dos seus erros podem ter sido fruto de distração, pois, ela possuía uma considerável maturidade conceitual. Neste caso, considera-se que o método probabilístico não foi eficiente ao medir sua aprendizagem;
- Aluna 5: novamente julgou-se que a evolução da aluna fez jus a uma maior valorização concedida pelo método probabilístico;
- Aluno 6: ficou evidente pela entrevista realizada que este aluno respondeu a todas as questões de modo aleatório (eram de múltipla escolha). Logo, o procedimento probabilístico, utilizando pesos, acertadamente reduziu sua nota, comparado a prova tradicional;
- Aluna 7: extra-sala essa aluna apresentou excelente desempenho, participando de atividades de monitoria e buscando aprender por outros meios. Deste modo, considerou-se justo o aumento de nota concedido pelo método;

- Aluna 8: esta aluna mostrou-se não participativa ao longo de todo o curso e, observou-se pela entrevista, que alguma evidência de fraude ou resposta aleatória existia, pois, nem todos conceitos estavam bem formulados. Deste modo, entendeu-se que o método reduziu justamente a nota final;
- Aluno 9: assim como em alguns casos apresentados este aluno mostrou elevada evolução do início ao fim do curso, sendo justa a melhoria da nota;
- Aluna 10: esta aluna, quando realizou o teste de habilidade inicial antes do período de estudo, o fez com extremo descaso, fato que gerou uma habilidade muito baixa. Logo, o método probabilístico considerou equivocadamente que houve uma grande evolução e propiciou um ligeiro e indevido aumento na sua nota;
- Aluna 11, 12, 13 e 14: todos estes alunos apresentaram evolução conceitual elevada, fazendo também jus à melhoria da nota que o método testado lhes atribuiu.

Deste modo, constatou-se que na maioria de 12 ocorrências a avaliação informatizada probabilística mostrou-se mais eficaz do que o procedimento tradicional e em apenas 2 casos (Alunos 4 e 10) ela foi considerada inadequada.

Aliado a estes resultados, Sands e Waters (1997, apud COSTA, 2009) apresentam outros argumentos quanto às vantagens e desvantagens de testes informatizados.

Vantagens:

- Redução do tamanho do teste;
- Nenhum item administrado no teste é irrelevante, pois, cada item apresentado ao aluno é adequado de acordo com sua habilidade;
- Flexibilidade para realizar baterias de testes, pois, o examinando não precisa aguardar todos acabarem para passar para a prova seguinte;
- O administrador do teste pode fornecer as instruções do teste virtualmente, sem precisar repetir para cada aluno que chegar atrasado e com isso não atrapalha a concentração dos examinandos que já tiverem começado o teste;
- Maior austeridade no controle das regras do teste, estando menos sujeito à burla de regras como a do tempo-limite para a realização do teste;
- Simplificação no processo de correção do teste;
- Resultados sendo publicados imediatamente após sua realização;
- Redução de erros ocorridos em processo de correção;
- Ajustam adequadamente o nível de dificuldade das questões aos examinandos, sem danificar a apuração das estimativas;
- Os testes adaptativos informatizados possibilitam a utilização de itens com recursos multimídia, o qual torna mais interesse aos alunos do que os testes tradicionais;
- Tempo de resposta a um item, permitindo ao avaliador saber o tempo que o aluno gastou em seu desenvolvimento.

Desvantagens:

- Requerem recursos financeiros para informatização de ambientes e capacitação humana para a sua construção e administração;
- Embora a segurança dos testes informatizados seja uma das vantagens, ela também pode se tornar um de seus maiores problemas caso os sistemas utilizados não sejam de alta confiabilidade (o que também implica em altos investimentos);
- Banco de itens deve ser continuamente atualizado para a segurança dos testes.

5 CONSIDERAÇÕES FINAIS

Embora seja uma tecnologia amplamente utilizada no exterior, os testes adaptativos informatizados ainda são pouco usados no Brasil. Neste trabalho, deu-se destaque a dissertação de Oliveira (2002) a tese de Vieira Junior (2012), na qual se desenvolveu o *software* educacional Sophia (usado nessa pesquisa), que também atua como sistema informatizado de avaliação. Embora, existam também outras iniciativas na literatura, ainda é preciso maior investimento e desenvolvimento nesta área.

Fora isto, discute-se também algumas das falhas promovidas por este método como as apontadas na análise de dados. Por exemplo, quando um aluno com habilidade inicial muito baixa poderia apresentar grande evolução fato que geraria uma média alta, mesmo que ele ainda não possuísse domínio de conteúdo em alto nível. Ainda assim acredita-se que um “erro” deste tipo seria menos significativo do que a somatória de todos os erros que já se observam, por exemplo, na própria avaliação tradicional. Por esta ótica, acredita-se que a informática é ainda um campo que tem muito a oferecer para auxiliar nos processos de ensino e, porque não, também nos processos de avaliação.

Referências

CATTAI, Adriano Pedreira. **Software matemático**. Bahia, 2009. Disponível em: <http://cattai.mat.br/site/files/sm/SoftwareMatematico_Cattai.pdf>. Acesso em: 26 ago. 2015.

HAMBLETON, R. K.; SWAMINATHAN, H. **Item response theory: principles and applications**. Boston: Kluwer Nijhoff Publishing, 1985. In: OLIVEIRA, Leandro Henrique Mendonça de. **Testes adaptativos sensíveis ao conteúdo do banco de itens: uma aplicação em exames de proficiência em inglês para programas de pós-graduação**. 2002. Dissertação (Mestrado em Ciências da Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2002.

Niltom. **Planejamento de um ambiente virtual de aprendizagem baseado em interfaces dinâmicas e uma aplicação ao estudo de potência elétrica**. 2012. Tese (Doutorado em Engenharia Elétrica) - Faculdade de engenharia de Ilha Solteira, Universidade Estadual Paulista, UNESP, Ilha Solteira, 2012.

OLIVEIRA, Leandro Henrique Mendonça de. **Testes adaptativos sensíveis ao conteúdo do banco de itens: uma aplicação em exames de proficiência em inglês para programas de pós-graduação**. 2002. Dissertação (Mestrado em Ciências da Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2002.

SANTOS, D. F.; RIBEIRO, L. G.; GUEDES, L. G. R.; MARTINS, W. Ferramenta avaliativa pedagógica para cursos a distância baseada em testes adaptativos informatizados e teoria de resposta ao item. In: SEMINÁRIO DE COMPUTAÇÃO, 14., 2004, Blumenau. **Anais...** Blumenau: Sociedade Brasileira de Computação, 2004.

SANDS, W. A.; WATERS, B. K. (Ed.). **Introduction to ASVAB and CAT**. Em Computerized Adaptive Testing: from inquiry to operation. Washington: American Psychological Association, 1997. In: COSTA, Denise Reis. **Métodos Estatísticos em Testes Adaptativos Informatizados**. 2009. Dissertação (Mestrado em Estatística) – Departamento de Métodos Estatísticos da Universidade Federal do Rio de Janeiro, UFRJ, Rio de Janeiro, 2009. Disponível em: <<http://www.pg.im.ufrj.br/teses/Estatistica/Mestrado/121.pdf>>. Acesso em: 21 ago. 2015.

