

## TESTES ADAPTATIVOS

Agostinho Iaquan Ryokiti Homa y Claudia Lisete Oliveira Groenwald  
Universidade Luterana do Brasil  
iaqchan@hotmail.com, claudiag@ulbra.br

Brasil

**Resumen.** O objetivo desse trabalho foi desenvolver um Teste Adaptativo Computacional (TAC) para conteúdos matemáticos, dentro do padrão SCORM (Sharable Content Object Reference Model). Um TAC é um sistema que apresenta questões com níveis de dificuldades adequadas ao desempenho do estudante, com a intenção de identificar as habilidades e competências do mesmo. Para isso é necessário um banco de questões com uma boa qualidade psicométrica e um algoritmo de seleção das questões a serem apresentadas. Do ponto de vista psicométrico, a Teoria de Resposta ao Item (TRI) ofereceu suporte ao desenvolvimento do algoritmo para o TAC proposto. A adoção, como premissa básica, do padrão SCORM possibilita que o teste desenvolvido possa ser utilizado em qualquer ambiente de aprendizagem eletrônica compatível com o padrão.

**Palabras clave:** tecnologias na educação, TAC, SCORM

**Abstract.** The aim of this study was to develop a Computer Adaptive Test (CAT) for mathematical content within the standard SCORM (Sharable Content Object Reference Model). A CAT is a system that presents questions with difficulty levels suitable to the student's performance, with the intention of identifying the skills and competencies. This requires a bank of questions with good psychometric quality and a selection algorithm of the issues to be presented. From the psychometric perspective, the Item Response Theory (IRT) offered support to the development of the algorithm for the proposed CAT. The adoption, as a basic premise, of the SCORM standard enables that the developed test can be used in any electronic learning environment that supports the standard

**Key words:** technologies in education, CAT, SCORM

### Introdução

Testes Adaptativos Computacionais, como: Graduate Record Examination (GRE); Test of English as a Foreign Language (TOEFL), desenvolvidos pela Educational Testing Service (ETS), e Armed Services Vocational Aptitude Test Battery (ASVAB), foram desenvolvidos e implementados em plataformas dedicadas, isto é, que utilizam sistemas configurados especialmente para atender necessidades e funcionalidades propostas. As configurações diferenciadas, tanto nos modos de apresentação quanto na formatação de suas bases de dados, impossibilitam o intercâmbio de conteúdos entre plataformas dedicadas de forma simples e funcional.

Para a troca de informações, entre plataformas dedicadas de diferentes fabricantes ou instituições e sem padronização, é necessária a construção de ferramentas para conversão e importação, ou que se faça a implementação, desde o início, na plataforma de avaliação a ser utilizada. Dessa maneira não há portabilidade, isto é, que os conteúdos, com um objetivo instrucional específico, tenham sempre o mesmo comportamento de apresentação independente da plataforma utilizada, podendo ser importados por qualquer sistema de gerenciamento de aprendizagem (*Learning Management System - LMS*).

Nesse sentido o SCORM pela *Advanced Distributed Learning Initiative* (ADLNET, 2009) define um padrão para o *e-learning*.

### Problema de pesquisa

Atualmente, testes em módulos instrucionais segundo o padrão SCORM, apresentam as questões em sequências definidas pelo autor, ou de forma aleatória, apresentando quaisquer questões que compõem o banco de questões disponível. Nesse sentido, surge o problema dessa investigação: como desenvolver um Teste Adaptativo Computacional, para uso em sequências de aprendizagem desenvolvidas segundo o padrão SCORM?

### Objetivo da

O objetivo desse trabalho foi desenvolver uma alternativa para o acompanhamento da aprendizagem, identificando as competências e habilidades dos estudantes através do uso de um Teste Adaptativo Computacional (TAC), fundamentado no modelo probabilístico da Teoria de Resposta ao Item (TRI), aplicado em *e-learning*, desenvolvidos segundo o padrão SCORM, com conteúdos matemáticos.

### Metologia da investigação

As ações de pesquisa foram divididas em duas etapas: estudo de literaturas para embasamento teórico sobre os TAC e a teoria que os fundamentam; desenvolvimento de um protótipo, investigando as funcionalidades através de comparações entre as habilidades estimadas pela ferramenta e as habilidades adotadas em diferentes simulações de respostas.

### SCORM

A *Advanced Distributed Learning Initiative* (ADL) foi criada em 1997 para padronizar e modernizar os processos de formação e educação sendo supervisionado pelo Departamento de Defesa (DoD) dos Estado Unidos da América. O ADL é um esforço colaborativo para aproveitar o poder das tecnologias da informação para oferecer uma educação de alta qualidade, acessível, adaptável e rentável. Formado por um grupo multinacional, composto de colaboradores do ramo da indústria, da academia e governo, que ajudam a definir as especificações e padrões para *e-learning*. Este grupo desenvolveu o *Sharable Content Object Reference Model* (SCORM®).

O SCORM é o resultado da união de especificações e padrões, das diversas entidades normativas, que definem as inter-relações entre conteúdos de aprendizagem, modelos de dados, protocolos e como os conteúdos são compartilhados entre sistemas de *e-learning* que seguem o mesmo padrão na mesma versão, sua edição mais atual é o *SCORM 2004 4rd Edition*.

O SCORM é um documento técnico que especifica regras e padrões, para os criadores de ferramentas de autoria, com definições de comportamento para apresentação dos conteúdos de aprendizagem. O padrão não foi elaborado diretamente para desenvolvedores ou designers de conteúdos, pois não define como criar lógicas instrucionais ou melhores práticas de ensino. As definições se restringem aos diversos pontos de análise e decisão para determinar o comportamento da plataforma relativo à apresentação dos conteúdos ao aluno.

### Teoria de resposta ao item TRI

Pela impossibilidade de medição direta desenvolveram-se Técnicas de Psicometria, que levaram às definições de mensuração, escala, ordenação e classificação das medidas de variáveis latentes (Tenant, 2004; Pallant & Tenant, 2007; Cicchetti et al, 2006 apud Chachamovich, 2007).

Uma teoria de grande destaque na Psicometria é a Teoria de Rasch, pois proporciona a derivação de medidas quantitativas, intervalares e aditivas ao colocar itens (item se refere à questão do teste) e respondentes em uma mesma escala (Hambleton & Swaminathan, 1985 apud Chachamovich, 2007).

Um teste adaptativo, baseado na TRI, requer um banco com questões capazes de identificar respondentes com habilidades diferenciadas, ou seja, discriminar um respondente com uma habilidade de outro respondente que tenha uma habilidade imediatamente superior, ou inferior, na escala adotada. Tal característica é denominada *Discriminação do Item*.

A avaliação não deve ser passível de ser influenciada por conhecimentos característicos de grupos específicos de respondentes, ou seja, a probabilidade de acerto dos itens deve ser a mesma, independente de sexo, grupo socioeconômico, ou qualquer outra característica que diferencie os respondentes (Gamerman, Gonçalves & Soares, 2007), além da habilidade no campo de conhecimento que se esteja avaliando.

Na Teoria de Rasch, (Baker, 2001), a probabilidade de acerto  $P$  de um determinado item, com uma dificuldade  $d$ , por um respondente, com uma competência ou habilidade  $\theta$ , é dada por:

$$P = \frac{1}{1 + e^{-(\theta - d)}}$$
, ou seja, uma questão com um nível de dificuldade igual à habilidade do respondente terá uma probabilidade de 0,5 de acerto.

Com as pesquisas em Psicometria, o modelo de Rasch foi aprimorado para o denominado modelo probabilístico de dois parâmetros que, além da dificuldade do item, considera também o parâmetro de discriminação do item, ou a declividade da curva de discriminação, denominado por  $c$ . Testes com fatores de discriminação inferiores a 0,7 devem, para efeitos práticos, ser

desconsiderados, pois não agregam precisão às estimativas de habilidade. Logo o modelo probabilístico fica determinado por  $P = \frac{1}{1+e^{-c(\theta-d)}}$ .

Os modelos probabilísticos, da TRI, de um e dois parâmetros consideram a habilidade do respondente como único fator para a resolução do item. Mas em testes computacionais é comum o uso de testes de múltipla escolha, logo, a probabilidade de acerto por um respondente sem habilidades ou competências para um dado teste será de  $p$ , para  $p = \frac{1}{n}$ , onde  $n$  é o número de opções disponíveis, pois há a possibilidade de acerto casual mesmo que a habilidade do respondente seja muito inferior à dificuldade da questão apresentada. O modelo de três parâmetros é dado por:  $P = p + (1 - p) \frac{1}{1+e^{-c(\theta-d)}}$ .

### ○ teste adaptativo computacional desenvolvido

Um Teste Adaptativo Computacional é um sistema que apresenta, como um examinador humano faria, questões com níveis de dificuldades adequadas ao desempenho do estudante, com a intenção de identificar as habilidades e competências do conhecimento (Costa, 2009). O TAC desenvolvido tem como característica principal, atender a portabilidade descrita no padrão SCORM, para tal, foi adotado o padrão XML (Extensible Markup Language), seguindo uma ontologia própria, mas flexível e expansível. A portabilidade do XML atende as definições do SCORM, pois sendo um arquivo de texto, o mesmo pode ser incluído no Pacote de Conteúdos na forma de um arquivo comprimido, no formato PKZIP v2.04g, podendo ser importado para qualquer plataforma que adote o padrão.

O arquivo XML, nessa versão, suporta dois modos de apresentação, resposta de múltipla escolha e resposta discreta e por ser um arquivo texto, para a inclusão/exclusão de questões é necessário somente incluir/excluir o bloco de chave *Question* referente ao item. Para inclusão/alteração o atributo *level* tem valor, entre 1.0 e 5.0 e correspondente à dificuldade do item; a chave *Text* é o enunciado; a chave *Answer* a resposta certa e as erradas, onde há somente uma resposta certa com o atributo *right* igual a *true*, o texto auxiliar que será apresentado ao aluno, quando o TAC é iniciado em modo exercício e o mesmo responder incorretamente à questão, é definido na chave *Textwrong* e o apontamento para o arquivo de imagem auxiliar, caso haja, a ser apresentado juntamente com a questão, na chave *Image*. As imagens auxiliares (gráficos, diagramas, etc...) são automaticamente redimensionadas e apresentadas em uma resolução de 400 x 300 pixels, sendo necessários cuidados para que textos inclusos nos gráficos não fiquem ilegíveis.

A ferramenta foi desenvolvida em Adobe Flash e programação *actionscript 2.0*, pela característica de seus conteúdos não serem editáveis, isto é, seu programa não pode ser alterado, preservando

o trabalho intelectual de programação, assim como evita fraudes por parte dos respondentes, caso os mesmos usem de técnicas de programação reversa para descobrir as respostas corretas.

Nesse trabalho, o algoritmo de seleção para a escolha do próximo item é feita pela estimativa da proficiência do respondente para o próximo item, que também considera os itens anteriores, usando o procedimento iterativo de máxima verossimilhança (Baker, 2001) dado por

$$\theta_{s+1} = \theta_s + \frac{\sum_{i=1}^n c_i [u_i - P_i(\theta_s)]}{\sum_{i=1}^n c_i^2 P_i(\theta_s) Q_i(\theta_s)}, \text{ ou seja, } \theta_{s+1} = \theta_s + \Delta\theta_s \text{ onde:}$$

$\theta_{s+1}$  é a estimativa da habilidade do respondente para o próximo item;

$\theta_s$  é estimativa da habilidade do respondente para o item atual para  $s$  interações;

$c_i$  é o fator de discriminação do item  $i$ ,  $i = 1, 2, 3 \dots n$ ;

$u_i$  é a resposta do respondente para o item  $i$ , sendo considerado 1 para uma resposta certa e 0 para uma resposta errada;

$P_i(\theta_s)$  é a probabilidade de acerto do item  $i$  considerando o respectivo  $c_i$  do item e a estimativa da habilidade  $\theta$  com  $s$  interações, sendo determinado pelos modelos de Rasch;

$Q_i(\theta_s)$  é a probabilidade de erro do item  $i$  dado por  $1 - P_i(\theta_s)$ .

$\Delta\theta_s$  é o fator de correção da estimativa da habilidade do respondente para  $s$  interações e dado pela divisão dos somatórios,  $\frac{\sum_{i=1}^n c_i [u_i - P_i(\theta_s)]}{\sum_{i=1}^n c_i^2 P_i(\theta_s) Q_i(\theta_s)}$ .

A probabilidade de acerto do item seguiu, nesse trabalho, o modelo de dois parâmetros,  $P = \frac{1}{1+e^{-c(\theta-d)}}$ , para as questões discretas e  $P = p + (1-p) \frac{1}{1+e^{-c(\theta-d)}}$ , para as questões com resposta de múltipla.

Para a pesquisa foi gerado um banco de questões com dificuldades entre um e cinco, sendo a dificuldade três, considerada mediana. Para o cálculo das probabilidades e a máxima verossimilhança os valores são convertidos, internamente no programa, para uma escala de menos três a três, com zero sendo a dificuldade mediana. O parâmetro de discriminação do item foi considerado com valor unitário,  $c = 1$ , dessa maneira o modelo probabilístico para questões discretas considera somente a dificuldade do item ( $\theta$ ), e para as questões com resposta de múltipla escolha é considerada a dificuldade do item, ( $\theta$ ), e acerto casual do item  $p = \frac{1}{n}$ , para  $n$  o número de respostas apresentadas.

O TAC considera, *a priori*, uma habilidade mediana,  $\theta_1 = 0$ , apresentando primeiramente um item compatível com uma dificuldade mediana,  $d_1 = 0$ .

O algoritmo, adotado para a escolha do item, realiza o procedimento de máxima verossimilhança e determina a dificuldade do próximo item a ser apresentado ( $d_{s+1}$ ) arredondando o valor da estimativa da habilidade do respondente para o próximo item ( $\theta_{s+1}$ ), e associando o valor diretamente à dificuldade do próximo item ( $d_{s+1} = \text{Arredondamento}(\theta_{s+1})$ ), pois de acordo com a TRI, a habilidade e a dificuldade, devem estar em uma mesma escala.

O TAC desenvolvido termina o teste quando o fator de correção for tão pequeno que o valor da estimativa da habilidade do respondente ( $\theta$ ), se estabiliza e se tem um determinado grau de certeza na estimativa, ou seja, um Erro Padrão (EP) pequeno.

Segundo Baker (2001), o erro padrão (EP) na estimativa da habilidade fornece uma indicação da precisão estimada. O erro padrão é uma variação de valores em torno do parâmetro de habilidade do respondente podendo ser computada usando a equação  $EP = \frac{1}{\sqrt{\sum_{i=1}^n c_i^2 P(\theta) Q(\theta)}}$ .

Para o critério de parada foram definidos,  $\Delta\theta \leq 0,009$  e  $EP \leq 0,9$ , como parâmetros de satisfação para que o TAC interrompa a apresentação dos itens.

Com a intenção de focar o TAC com conteúdos de Matemática, explorando a possibilidade de respostas discretas com a possibilidade mínima de acerto casual, foram feitos testes simulados com itens de resposta discreta com probabilidades de acerto dado por  $P(\theta) = \frac{1}{1+e^{-(\theta-d)}}$  e uma estimativa da habilidade inicial do respondente dada por  $\theta_1 = 0$ .

As tabelas 1, 2 e 3 apresentam testes simulados para respondentes com habilidades  $\theta_1 = 2,0$ ;  $\theta_2 = 0,5$  e  $\theta_3 = -1,5$  respectivamente. Após a satisfação dos critérios de parada,  $\Delta\theta \leq 0,009$  e  $EP \leq 0,9$ , foram estimadas as respectivas habilidades  $\theta_1 = 2,243$ ;  $\theta_2 = 0,841$  e  $\theta_3 = -1,758$  para os respondentes, que se aproximam bem das habilidades simuladas.

Item	Estimativa da Habilidade	Dificuldade do Item	Fator de correção	EP	Resposta certa
1	0	0	-	-	1
2	1,000	1	1,000	2,000	1
3	2,890	3	1,512	1,577	0

4	2,244	2	0,057	1,439	0
5	1,500	2	0,000	1,141	1
6	2,120	2	0,000	1,007	1
7	2,512	3	0,000	0,930	0
8	2,243		0,000	0,835	

Tabela 1: Teste simulado 1, com habilidade do respondente estimada em 2,0

Item	Habilidade	Dificuldade do Item	Fator de correção	EP	Resposta certa
1	0	0	-	-	1
2	1,000	1	1,000	2,000	0
3	0,502	1	0,243	1,467	1
4	1,401	1	0,024	1,247	0
5	0,758	1	0,000	1,023	1
6	1,228	1	0,000	0,928	0
7	0,841		0,000	0,830	

Tabela 2: Teste simulado 2, com habilidade do respondente estimada em 0,5

Item	Habilidade	Dificuldade do Item	Fator de correção	EP	Resposta certa
1	0	0	-	-	0
2	-1,000	-1	-1,000	2,000	1
3	-0,502	-1	-0,243	1,467	0
4	-1,401	-1	-0,024	1,247	0
5	-1,896	-2	-0,002	1,169	0
6	-2,501	-3	0,000	1,153	1
7	-2,140	-2	0,000	0,952	1
8	-1,758		0,000	0,83	

Tabela 3: Teste simulado 3, com habilidade do respondente estimada em -1,5

Para os casos de erro ou acerto de todos os itens, o procedimento de máxima verossimilhança não atende os critérios de parada, sendo necessário que o algoritmo trate esses casos, limitando os valores estimados das habilidades do respondente ( $\theta_{s+1}$ ), tratando a não convergência do fator de correção para valores inferiores ao critério de parada,  $\Delta\theta \leq 0,009$ . Nesses casos o critério de parada é definido pela verificação de estabilidade na estimativa da habilidade do respondente em seus valores, máximo ou mínimo, em conjunto com o parâmetro de  $EP < 0,9$ .

Foram feitos testes simulados, para validação do tratamento e do critério de parada do algoritmo do TAC desenvolvido, nos casos de acerto e erro de todos os itens, a Tabela 4 apresenta a simulação de acerto de todos os itens.

Item	Habilidade	Dificuldade do Item	Fator de correção	EP	Respostas
1	0	0	-	-	1
2	1,000	1	1,000	2,000	1
3	2,890	3	1,512	1,577	1
4	3,000	3	1,666	1,581	1
5	3,000	3	1,794	1,240	1
6	3,000	3	1,851	1,054	1
7	3,000	3	1,884	0,932	1
8	3,000		1,905	0,845	

Tabela 4: Teste simulado com acerto de todos os itens

### Conclusão

A mudança do paradigma baseado na Teoria Clássica do Teste com escore bruto, para um comportamento adaptativo com a apresentação de questões compatíveis com as habilidades estimadas dos respondentes em um processo interativo durante todo o teste, segundo as técnicas de Psicometria da Teoria de Resposta ao Item, e adotando o padrão SCORM para definição dos seus requisitos de portabilidade e acessibilidade, mostrou ser mais condizente para a estimativa da habilidade do respondente assim como de fácil reutilização em outras sequências didáticas.

A adoção dos padrões de apresentação em *HTML* e banco de questões em *XML* atende, em acessibilidade e portabilidade, ao padrão SCORM. O padrão XML também se mostrou bem prático para inclusão/exclusão de questões por *designers* instrucionais sem a necessidade de conhecimentos específicos de informática.

Na simulação 2 o comportamento de respostas foi mais estável, no qual as respostas ficaram oscilando entre dois níveis de dificuldade, com a ferramenta estimando uma habilidade intermediária e compatível com as dificuldades dos itens apresentados. Nas simulações 1 e 3, foram apresentadas as situações onde as respostas oscilaram mais e de acordo com as possíveis probabilidades de erro e acerto para as habilidades adotadas no início do teste, e o algoritmo do TAC estimou com as mesmas com considerável precisão.

Tanto nas simulações de habilidade extremas, acerto e erro de todos os itens, como nas demais simulações, o algoritmo do TAC estimou a habilidade do respondente condizente com as habilidades adotadas para responder aos testes.

O TAC desenvolvido já foi avaliado em um grupo de alunos, do ensino superior, com o conteúdo de Análise Combinatória, onde apresentou todas as funcionalidades propostas e atingiu os objetivos propostos. O experimento está disponível em <http://matematica.ulbra.br/ilias>.

Com este trabalho, pretende-se a divulgação, em meio acadêmico da área de Educação Matemática, do uso de testes adaptativos no padrão SCORM em plataformas de ensino e também, a discussão e reflexão do acompanhamento da aprendizagem através de testes que se adaptam às habilidades dos respondentes, permitindo ao avaliador uma maior precisão das habilidades e aprendizagens conseguidas pelos alunos.

### Referências bibliográficas

Armed Services Vocational Aptitude Test Battery. (2010). Recuperado em janeiro de 2010 em: <http://www.military.com/ASVAB>

Advanced Distributed Learning Initiative. (2006). *SCORM 2004 3rd Edition Content Aggregation model*. Recuperado em março de 2010 em: <http://www.adlnet.gov/resources#Tab-Technical-Documentation>

Advanced Distributed Learning Initiative. (2006). *SCORM 2004 3rd Edition Sequencing and Navigation*. Recuperado em março de 2010 em: <http://www.adlnet.gov/resources#Tab-Technical-Documentation>

Advanced Distributed Learning Initiative. (2006). *SCORM 2004 3rd Edition Run-Time Environment*. Recuperado em março de 2010 em: <http://www.adlnet.gov/resources#Tab-Technical-Documentation>

Baker, F. B. (2001). *The basics of item response theory*. (2.ed.) Washington DC: Eric.

Chachamovich, E. (2007). *Teoria de resposta ao item: aplicação do modelo Rasch em desenvolvimento e validação de instrumentos em saúde mental*. Tese de Doutorado Programa de Pós-Graduação em Ciências Médicas: Psiquiatria, Faculdade de Medicina, Universidade Federal do Rio Grande do Sul, Porto Alegre.

Costa, D. R. (2009). *Métodos estatísticos em testes adaptativos informatizados*. 2009. 107 f. Dissertação de Mestrado. Instituto de Matemática, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

Gamerman, D., Gonçalves, F. B., & Soares, T. M. (2007). Análise bayesiana do funcionamento diferencial do item. *Pesquisa Operacional*, 27(2), 271-291.

Graduate Record Examination. (2010). Recuperado em janeiro de 2010 em:  
<http://gradschool.about.com/cs/aboutthegre/f/grefaq.htm>

Test of English as a Foreign Language. (2010). Recuperado em janeiro de 2010 em:  
<http://www.ets.org/toefl/>.