
Métodos de avaliação informatizada que tratam o conhecimento parcial do aluno e geram provas individualizadas

Jean Piton-Gonçalves¹²³, Andrea Jessica Borges Monzón², Sandra Maria Aluísio¹²

¹Instituto de Ciências Matemáticas e de Computação (ICMC)
Universidade de São Paulo (USP)
Caixa Postal: 668 – CEP: 13560-970 – São Carlos – Brazil

²Núcleo Interinstitucional de Linguística Computacional (NILC) – ICMC/USP

³Departamento de Matemática (DM)
Universidade Federal de São Carlos (UFSCar) – São Carlos – Brazil.

jpiton@ufscar.br, monzon.andrea@hotmail.com, sandra@icmc.usp.br

Abstract. *This paper presents two alternative computer assessment methods for large scale evaluation applied to graduate students from a master's degree program: the Admissible Probability Measurement and the Computer Adaptive Test based on the algorithm CBAT-2 as well as the Item Response Theory. The computational environments called CAPTEAP and CALEAP-Web show that alternative computer assessment methods may have advantages compared to traditional methods. Results from several years of application of them reinforce the hypothesis that it is possible having computational environments which focus on measuring the partial knowledge of students of English for specific purposes and on generating customized tests.*

Resumo. *O presente artigo apresenta dois métodos alternativos de avaliação informatizada em larga escala aplicados para os estudantes de um programa de mestrado: a Medida de Probabilidade Admissível e o Teste Adaptativo Informatizado baseado no algoritmo CBAT-2 e na Teoria de Resposta ao Item. Os ambientes computacionais denominados de CAPTEAP e CALEAP-Web mostram que os métodos informatizados alternativos podem apresentar vantagens em relação aos métodos convencionais. Resultados de vários anos de aplicações reforçam a hipótese de que é possível ter-se ambientes computacionais focados na medida de conhecimento parcial de estudantes de língua inglesa para propósitos acadêmicos e na geração de testes personalizados.*

Palavras-chave: Avaliação Educacional Informatizada, Medida de Probabilidade Admissível, Teste Adaptativo Informatizado.

1. Introdução

Nas últimas décadas, a preocupação com a qualidade do processo educacional nos contextos escolar e universitário tem se acentuado, e como parte desse processo, a avaliação dos estudantes. Isso é resultado das novas necessidades com as quais a sociedade do conhecimento vem se adaptando, com demandas de novos conhecimentos para o trabalho e

estudo, e a resignificação de conhecimentos já consolidados. Em muitos casos a avaliação do conhecimento do estudante é um processo que determina os rumos de sua vida, como acontece, por exemplo, em vestibulares, exames de proficiência e concursos públicos. Esses cenários são caracterizados por um processo de Avaliação Educacional em Larga Escala (AveLE). São exemplos no Brasil o vestibular da FUVEST¹, que em 2009 obteve 138.242 candidatos inscritos, e o Exame Nacional do Ensino Médio (ENEM) que na sua primeira edição, em 1998, contou com cerca de 157.000 estudantes, e na última, em 2008, contou com mais de 1 milhão de candidatos inscritos.

Esses números mostram que a demanda por avaliar estudantes em larga escala vem crescendo ao longo dos anos. A questão que permeia esse processo é: “*como avaliar em larga escala mantendo a qualidade, rapidez e equidade?*”. Como exemplos de resposta para a questão colocada acima, o presente artigo destaca dois métodos de avaliação educacional viabilizados por meio dos ambientes computacionais *Computer Assisted Proficiency Test of English for Academic Purpose (CAPTEAP)* e *Computer-Aided Learning English for Academic Purposes (CALEAP-Web)*, aplicados em cenários reais de avaliação.

A Seção 2 apresenta conceitos sobre avaliação, uma classificação para ela, além de trazer os tipos de testes objetivos e seus métodos convencionais e alternativos. As Seções 3.1 e 3.2 tratam dos métodos informatizados alternativos. A Seção 3.3 apresenta a Teoria de Resposta ao Item (TRI) que ampara a seleção e a calibração das questões a serem aplicadas aos alunos em testes informatizados. A Seção 4 descreve as características do teste usado no Exame de Proficiência em Inglês (EPI) do Programa de Pós-Graduação do Instituto de Ciências Matemáticas e de Computação (ICMC) da Universidade de São Paulo (USP), e do ambiente computacional de aprendizagem que prepara o estudante para o teste formal. A Seção 5 mostra a facilidade e maior adequação da calibração de novas questões sob o quesito dificuldade quando utiliza-se modelos da TRI.

2. Avaliação educacional

Testar, avaliar e medir são conceitos que podem ser confundidos entre si. Testar é o mesmo que verificar um desempenho mediante situações previamente organizadas; avaliar é um processo de interpretação de dados quantitativos e qualitativos com o objetivo de se obter um parecer ou julgamento de valor, tendo por base padrões ou critérios; medir descreve um fenômeno do ponto de vista quantitativo, o qual tem por base um sistema de medidas [Haydt 1988].

A avaliação educacional pode ser classificada segundo [Miller et al. 1998]: (a) *Avaliação Diagnóstica*, que verifica a presença ou ausência de pré-requisitos necessários para o estudante inserir-se em um novo contexto de aprendizado; (b) *Avaliação Formativa*, que contribui para o aprendizado do estudante durante o processo, tendo como função acompanhar sua aprendizagem; e (c) *Avaliação Somativa ou Final* caracterizada por ser realizada após um ciclo de estudos, visando a promoção de um nível para outro ou a classificação do estudante segundo uma determinada escala.

Na AveLE caracterizada pela avaliação somativa, as questões podem ter resposta dissertativa ou objetiva. A primeira opção é definida pela argumentação escrita, necessitando de grandes grupos de pessoas especialistas em correções, que em determinados

¹Fundação Universitária para o Vestibular

cenários de larga escala pode ser custoso. Como alternativa, os modelos não tradicionais de avaliação por resposta objetiva têm se mostrado cada vez mais capazes avaliar com precisão os estudantes.

No contexto da avaliação objetiva, os testes objetivos requisitam do aluno uma resposta a uma questão cuja resposta correta é pré-determinada [McKenna and Bull 1999]. Os testes objetivos podem ser classificados segundo [Olea et al. 1999]: (a) *Teste Objetivo Tradicional* que tem a principal característica de realizar o teste via lápis e papel, com aplicação e correção realizadas por pessoas, manualmente; (b) *Teste Tipo Fichas* que são como os testes tradicionais, porém a correção é automatizada (por meio de leitores ópticos, por exemplo); (c) *Teste Objetivo Informatizado* que é totalmente automatizado, sendo que o computador tem a finalidade de fornecer as questões e os resultados relativos ao estudante por meio de tabelas, gráficos ou números, isto é, o computador aplica e corrige o teste.

Em um Teste Objetivo Informatizado, todo o processo é realizado pelo computador (exposição das questões, análise e resultados), podendo ser categorizado segundo dois critérios: forma de pontuação (certo/errado ou avaliação do conhecimento parcial) e elenco de questões (mesmo número e/ou mesmas questões e/ou variação no número de questões). Esses testes são agrupados em [Aquino 2001]:

Métodos Informatizados Convencionais: são aqueles que possuem o mesmo número de questões para todos os alunos participantes e cujas opções de resposta são certo/errado;

Métodos Informatizados Alternativos: medem o conhecimento parcial do aluno, ou geram quantidades diferentes de questões por teste, conforme a proficiência/habilidade do estudante, por exemplo.

3. Métodos Informatizados Alternativos

Um teste objetivo tradicional pode ser transposto para um teste informatizado, apresentando algumas vantagens tais como [Oliveira 2002] correção automatizada com produção de estatísticas, geração de relatórios individuais ou do grupo de alunos, monitoramento em tempo real de teste e controle do tempo de sessão do teste. Apesar disso, existe apenas a transposição do modelo tradicional para o informatizado, apresentando apenas vantagens sob o ponto de vista operacional, e com poucas vantagens educacionais.

Na perspectiva de ampliar as possibilidades de AvELE informatizada, serão apresentados neste artigo dois métodos informatizados alternativos: a Medida de Probabilidade Admissível [Shuford and Brown 1974, Bruno 1986, Klinger 1997] e o Teste Adaptativo Informatizado [Weiss and Kingsbury 1984, Olea et al. 1999]

3.1. Medida de Probabilidade Admissível

Um teste de múltipla escolha com opções de resposta dicotômicas apresenta dois sérios problemas pedagógicos: o “chute” e o pressuposto de que o conhecimento do aluno é “binário” (correto ou incorreto). Com o objetivo de solucionar parte destes problemas, a *Medida de Probabilidade Admissível* (MPA) [Shuford and Brown 1974] não incentiva o “chute” nas alternativas de resposta e permite avaliar o conhecimento parcial do aluno. Cada opção de resposta é interpretada por meio de um modelo que considera a escala de pontuação entre uma resposta correta ou incorreta, sendo gerada por uma probabilidade. A MPA informatizada será detalhada na Seção 4.1.

3.2. Teste Adaptativo Informatizado

Um *Teste Adaptativo Informatizado* (TAI) propõe uma forma de avaliação caracterizada por selecionar as questões segundo os níveis estimados de habilidade (ou proficiência) do estudante. Mediante um conjunto de questões, um TAI tem a principal característica de “individualizar um teste”, ou seja, cada aluno participante recebe um elenco de diferentes questões, em diferentes quantidades. Ainda, um TAI seleciona os itens de forma a maximizar a precisão do exame, baseando-se no conhecimento do aluno a partir do histórico das questões anteriormente respondidas [Weiss and Kingsbury 1984]. Desta forma, pode-se aplicar testes com maior flexibilidade e adaptabilidade, além da redução do tempo, resultados imediatos e maior precisão em relação ao Teste Objetivo Tradicional [Olea et al. 1999].

São elementos básicos de um TAI [Conejo et al. 2001]:

1. o *Modelo de Resposta ao Item*, que descreve como o aluno examinado responde ao item, segundo seu nível de habilidade estimado;
2. o *Método de Seleção dos Itens*, que proporciona a seleção do próximo item em função do nível estimado do conhecimento do aluno e da resposta do item previamente ministrado;
3. o *Banco de Itens (BI)*, que é um banco de dados composto pelos itens e outras informações inerentes à cada questão;
4. o *Nível de Conhecimento Inicial*, que envolve a escolha da primeira questão do teste;
5. o *Critério de Parada*, que é a determinação do término do teste, que pode ser o alcance de um nível pré-estabelecido de habilidade, quantidade máxima de itens em um teste, tempo de teste, etc;
6. a *Calibração de Itens*, que é a determinação dos parâmetros inerentes à questão.

Tratando-se da calibração dos itens do banco, ela pode ser por *Calibração Estática*, cujos parâmetros dos itens são pré-calibrados e recalibrados após cada teste, porém nunca durante a execução do teste (o modelo de calibração dos parâmetros provém da Teoria de Resposta ao Item); ou por *Calibração Dinâmica*, sendo exemplificada pelo *algoritmo CBAT-2* [Huang 1996], que elimina um estudo empírico para a calibração de itens do teste.

3.3. Teoria de Resposta ao Item

Como parte da elaboração de um TAI, a *Teoria de Resposta ao Item* (TRI) modela a probabilidade de um aluno responder a uma questão corretamente em função de sua habilidade (ou proficiência). Assim, quanto maior a proficiência, maior a probabilidade do aluno acertar a questão. A TRI considera uma família de modelos matemáticos que relacionam os itens (questões) de um teste com a habilidade estimada do estudante, e a teoria possui dois postulados básicos [Pasquali and Primi 2003]:

1. o desempenho do sujeito em uma questão pode ser predito a partir de um conjunto de fatores ou variáveis hipotéticas, ditos aptidões ou traços latentes; o θ sendo a causa e o desempenho o efeito;
2. a relação entre o desempenho e os traços latentes (habilidade) pode ser descrita pela Curva Característica do Item. Em linhas gerais, à medida que o θ aumenta, a probabilidade de acerto da questão também aumenta.

Um importante modelo de resposta ao item é o Modelo Logístico de Três Parâmetros (ML3P) [Lord 1980, Baker 2001], amplamente utilizado em cenários de AvELE. Como exemplos de exames que usam a TRI temos o SAEB², o SARESP³, o novo ENEM⁴ e o TOEFL⁵. Para o caso unidimensional para itens dicotômicos⁶ da TRI, a probabilidade do aluno com habilidade θ acertar a questão é dada por

$$P(\theta) = c + (1 - c) \frac{1}{1 + e^{-Da(\theta-b)}} \quad (1)$$

onde $P(\theta)$ é a probabilidade de um aluno com habilidade θ acertar a questão, a , com $a \in [0, \infty)$ é o parâmetro *Discriminação* que descreve como um item pode diferenciar os alunos de baixa habilidade dos alunos de alta habilidade; b , com $b \in (-\infty, \infty)$, é a *Dificuldade* de um item; c , com $c \in [0, 1]$, é o *chute* que representa a probabilidade de ter-se uma resposta correta mediante uma baixa habilidade estimada do examinado; e D é uma constante de ajuste da curva. A pré-calibração e a recalibração dos parâmetros a , b e c e a estimação de θ são realizadas por métodos específicos, descritos em Lord (1980).

Uma importante potencialidade da TRI é a possibilidade de comparar-se os resultados entre diferentes grupos de estudantes em diferentes níveis, mantendo sempre a equidade e a correlação no cálculo do escore dos estudantes.

4. Ambientes computacionais desenvolvidos

Partindo dos métodos alternativos informatizados, foram desenvolvidos dois ambientes computacionais que apóiam a avaliação e aprendizagem dos estudantes ingressantes no nosso programa de mestrado acadêmico quanto à proficiência em inglês. Eles serão abordados nas Seções 4.1 e 4.2.

4.1. CAPTEAP

Essencialmente a MPA classifica o conhecimento final do aluno como *bem informado*, *desinformado*, *mal informado*, *totalmente mal informado*, *não informado* e *parcialmente informado*, que possibilitam descrever o conhecimento ou desconhecimento sobre o conteúdo envolvido na questão [Bruno 1986]. Nesse contexto, foi desenvolvido o ambiente via Web CAPTEAP. Trata-se de um exame automatizado de proficiência em inglês para avaliar os estudantes ingressantes no Programa de Pós-Graduação da nossa instituição, que foi projetado e desenvolvido mediante a Taxonomia de Objetivos Educacionais de Bloom [Bloom and Krathwohl 1956], proporcionando a elaboração de questões de forma mais clara e a geração de exames equilibrados em termos de habilidades avaliadas [Aluísio et al. 2003].

No CAPTEAP, cada questão possui três alternativas de resposta propostas conforme mostradas na Figura 1. A escolha do estudante corresponde à seleção dentre as

²Sistema de Avaliação da Educação Básica - www.inep.gov.br/basica/saeb

³Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo - saresp.edunet.sp.gov.br

⁴Exame Nacional do Ensino Médio - www.enem.inep.gov.br

⁵Test of English as a Foreign Language - www.toefl.org

⁶São aqueles que trabalham apenas com uma única habilidade do aluno por teste, nos quais os itens a serem respondidos são dicotômicos, ou seja, dentre as opções de resposta existe apenas uma resposta correta.

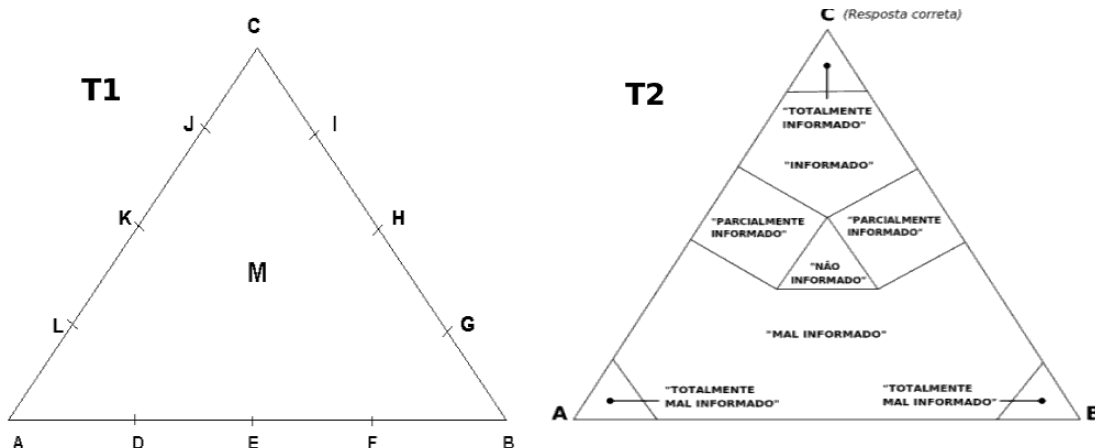


Figura 1. T1: Triângulo utilizado no ambiente. T2: Triângulo equilátero da MPA com a caracterização das probabilidades [Bruno 1986].

opções de resposta (de *A* à *M*) dispostas em um triângulo equilátero, em que cada resposta está associada a uma pequena região de probabilidades, representando a escolha de alguma alternativa. A Figura 1 ilustra três possíveis respostas em um teste de múltipla escolha, em que as alternativas *A* e *B* são incorretas e *C* é a alternativa correta. Assim, cada questão é pontuada com uma das classes relativas ao conhecimento do estudante: totalmente informado (letra *C*), quase informado (letra *J* ou *I*), parcialmente informado (letra *K* ou *H*), parcialmente mal informado (letra *G* ou *L*), mal informado (letra *A* ou *D* ou *E* ou *F* ou *B*) e desinformado (letra *M*). A Figura 2 mostra uma questão do CAPTEAP.

Part 4 : Question 12
 Texto: " MEASURE AND PROBABILITY FOR CONCURRENCY THEORISTS "
 Área: Geral
 Autor: Prakash Panangaden
 Fonte: Theoretical Computer Science 253 (2001) 287–309
 URL: www.elsevier.com/locate/tcs

Instructions

Contents

Part 1

Part 2

Part 3

Part 4

Questions

Part 5

Part 6

Part 7

EXIT !

In the area of semantics, one of the earliest serious investigations is due to Kozen [28, 29]. His work uses measuretheoretic ideas in a serious way and, in particular, uses conditional probability distributions or Markov kernels as a central tool. In his work he gives probabilistic semantics of a language of while loops and also develops a "Stonetype" duality using the idea that measurable functions are "probabilistic predicates". This analogue of "predicate transformer" semantics has been extensively developed by a group at Oxford [35]. In a different vein Gupta, Jagadeesan and Saraswat [18] have developed a modeling language for probabilistic systems based on the concurrent constraint programming paradigm. As soon as one adds recursion to the language [17] one is forced into the realm of continuous spaces and the ideas expounded in the present paper are relevant. The other main starting point for the use of probability theory in semantics was the work of Sahel/Djaroni [37, 38]. His work combined probability theory and domain theory and was the inspiration for probabilistic powerdomains [23, 22] by Jones and Plotkin. Ultimately this led to the enormously fruitful work of Edalat and others [12–14] on integration on domains. The work on verification of probabilistic systems has exploded in recent years - it is impossible to attempt a survey here. There are approaches based on automata theory, process algebra equivalences, logics and model checking. A very interesting development has been the use of a probabilistic process algebra as a compositional performance evaluation tool [20].

Which strategy follows in the example?

Options	Answers
A	Citations grouped by approach
B	Citations ordered from general to specific
C	Citations ordered chronologically

12 13

Figura 2. Tela do CAPTEAP apresentando uma questão.

O critério de aprovação no Exame de Proficiência em Inglês da nossa instituição é baseado em duas avaliações: Avaliação Global e Avaliação por Módulos. A aprovação de um estudante ocorre quando na Avaliação Global obtiver 50% (ou mais) de respostas na classe Totalmente Informado, e 25% (ou menos) de respostas na classe Mal Informado; ou 90% (ou mais) de respostas nas classes Totalmente Informado, Quase Informado e

Parcialmente Informado e 10% (ou menos) de respostas na classe Mal Informado.

Para os estudantes que se aproximarem da nota de corte da Avaliação Global, ou seja, aqueles com uma ou duas questões abaixo de 50%, ou uma ou duas acima de 25%, ou ainda uma abaixo de 50% e uma acima de 25%, são considerados aprovados se obtiverem, com relação ao módulo que trata da Estrutura de Textos Científicos, 50% (ou mais) de respostas na classe Totalmente Informado, 25% (ou menos) de respostas nas classes Mal Informado, Desinformado, Parcialmente Mal Informado; e por consequência 25% (ou mais) nas classes Quase Informado e Parcialmente Informado.

Tabela 1. Classificação promovida pela MPA no CAPTEAP.

Classificação	Significado	Estratégia de Ensino
Totalmente Informado	Conhecimento correto a respeito da questão	Avanço
Quase Informado	Conhecimento quase correto a respeito da questão	Avanço/Revisão
Parcialmente Informado	Conhecimento de uma parcela da questão	Revisão
Desinformado	Nenhum conhecimento a respeito da questão	Instrução
Parcialmente Mal Informado	Conhecimento próximo do errado a respeito da questão	Instrução/Re-educação
Mal Informado	Conhecimento errado a respeito da questão	Re-educação

O ambiente computacional mencionado fornece ao final do exame um relatório de desempenho do estudante, baseado na Tabela 4.1, com indicação explícita em qual critério houve aprovação no EPI de nossa instituição.

4.2. CALEAP-Web

Com o objetivo de apoiar o estudo do aluno de pós-graduação de nossa instituição sobre o gênero de textos científicos em inglês, e também prepará-lo para a avaliação formal de Proficiência em Inglês, foi desenvolvido o Ambiente Computacional de Aprendizagem CALEAP-Web. Tal ambiente foi projetado, desenvolvido e avaliado a partir da integração de um TAI baseado no CBAT-2 — que consiste em itens (questões) selecionados de acordo com o nível de habilidade estimado do aluno, gerando um teste individualizado com poucas questões — e um ambiente computacional que possui tarefas de aprendizagem elaboradas a partir das convenções do gênero de textos científicos.

Destacando o TAI do CALEAP-Web, este tem o papel de avaliar o aluno em diferentes habilidades linguísticas, distribuídas em quatro módulos dentro do conhecimento de inglês acadêmico: (1) *convenções da língua inglesa para artigos científicos*, (2) *estrutura esquemática de artigos científicos*, (3) *compreensão de texto* e (4) *estratégias de escrita do gênero em questão*.

A partir dos resultados de nossas pesquisas [Piton-Gonçalves et al. 2004, Piton-Gonçalves 2004, Aquino 2001] e um histórico de aplicação de quatro anos com usuários reais do CALEAP-Web – resultados da calibração e aplicação de 108 questões distribuídas nos quatro módulos do exame –, um TAI baseado no CBAT-2 apresenta as seguintes características:

1. A individualização de um teste, cujo conhecimento do examinado determina o seu próprio elenco de questões durante o teste, quando responde correta ou incorretamente a uma questão;
2. O tempo de teste em relação ao aluno é reduzido, mantendo o mesmo nível de confiança de um teste convencional. Isso é uma vantagem, pois reduz a fadiga que os longos testes causam nos alunos;

-
3. Pode-se utilizar o TAI como ferramenta para a Educação à Distância;
 4. O teste atinge um alto nível de precisão na avaliação do estudante com um baixo tempo de teste devido à adaptabilidade do TAI;
 5. Existem garantias de que um TAI baseado na TRI seleciona as questões que mais se aproximam da proficiência estimada do aluno;
 6. *Feedback* imediato ao estudante sobre seu desempenho no teste;
 7. Permite analisar e comparar equivalentemente os resultados de diferentes grupos de alunos a partir de diferentes questões;
 8. Um TAI baseado no CBAT-2 não necessita de uma pré-calibração, eliminando a aplicação de pré-testes. Nesse caso basta uma atribuição inicial pelo especialista do parâmetro dificuldade, tomando como base uma matriz de dificuldades que relaciona numericamente a dificuldade com uma escala inicial (aqui seria muito fácil, fácil, médio e difícil);
 9. A tendência do parâmetro dificuldade é estabilizar seu valor ao longo do tempo. Este valor de estabilidade é chamado de *Parâmetro b Real*, que expressa o conhecimento sobre a dificuldade do item a partir de todo o histórico de aplicação do item ao longo de todos os testes anteriormente aplicados;
 10. A calibração dos itens é realizada de acordo com dois tipos de conhecimento: histórico de aplicação dos itens e histórico das questões do teste imediatamente respondido pelo examinado;
 11. O TAI baseado no CBAT-2 e na TRI mostra-se como uma possível solução para instituições que possuem um banco de itens pequeno, e desejam obter o máximo de informação quanto à proficiência do estudante em determinado domínio de conhecimento.

Um TAI possui algumas limitações [Piton-Gonçalves 2004]:

1. O banco de itens de um TAI requer cuidados na elaboração, tanto em relação às questões quanto na sua calibração;
2. Por envolver métodos estatísticos, a implementação de um sistema baseado na TRI pode ser custosa e trabalhosa;
3. Deve-se implementar algoritmos que processem os cálculos de maneira segura e rápida, pois para cada questão respondida existe uma determinada complexidade de iterações numéricas.

5. Calibração de novas questões com a TRI

Pelo fato da nossa instituição visar o uso do inglês para fins específicos (ESP - English for Specific Purposes) [Hutchinson and Waters 1987] nas provas de proficiência, essas têm peculiaridades linguísticas relevantes que são devotadas à convenções da língua inglesa e estruturação de textos do gênero científico. Como o módulo 1 do EPI, que avalia convenções linguísticas em artigos científicos passou por alterações em seu conteúdo em 2008, houve a necessidade de que o CALEAP-Web a elas se adaptasse. Houve, portanto, a inserção de 51 questões de cunho gramatical primando pelo emprego de artigos definidos e indefinidos e de pronomes relativos. Para tanto, houve a contribuição de uma linguista que foi responsável pela criação de um banco de questões [Monzón 2008]. Além da elaboração dos itens em si, houve a preocupação de classificá-los por níveis de dificuldade - muito fácil, fácil, médio e difícil - e harmonizá-los em termos de quantidade.

Essas questões puderam ser calibradas enquanto os avaliandos se preparavam, através da aplicação do CALEAP-Web, para dois exames que foram aplicados em semestres diferentes a 54 e 13 avaliandos, respectivamente. No primeiro, houve a aprovação de 75% dos pós-graduandos e no segundo de 54%. Ao se analisar os resultados dos examinados, especificamente no módulo 1, nota-se que a mudança dos conteúdos não os afetou negativamente, ou seja, houve um bom rendimento em termos de acertos.

6. Considerações finais

Atualmente, há uma preocupação crescente com a qualidade do ensino, aprendizagem e avaliação nos mais diferentes níveis de ensino. Nessa perspectiva, este artigo busca explicar dois métodos alternativos de avaliação informatizada que pode ser aplicada em cenários de larga escala, desenvolvidos e avaliados em nossa instituição de ensino superior. Tanto o CAPTEAP quanto o CALEAP-Web, que são baseados na MPA e no TAI respectivamente, são ambientes computacionais utilizados pelos estudantes de mestrado do nosso programa de pós-graduação, para o estudo e avaliação quanto à proficiência em Inglês Acadêmico.

Os resultados obtidos, a partir dos trabalhos de pesquisadores da nossa instituição, reforçam a hipótese da possibilidade de ter-se uma AvELE informatizada que valoriza o conhecimento parcial do aluno e o conhecimento do grupo de alunos (possibilitado pelo CBAT-2 e TRI), que calibra dinamicamente o nível de dificuldade das questões a partir de um histórico de testes anteriormente aplicados e ter-se um banco de itens calibrado a partir do conhecimento dos alunos. Além disso, um teste informatizado baseado em métodos alternativos explora as potencialidades do computador, e não somente a transposição do sistema lápis e papel para o computador, em que há somente a automatização de um processo manual.

Referências

- Aluísio, S. M., de Aquino, V. T., Pizzirani, R., and Oliveira, O. N. (2003). High order skills with partial knowledge evaluation: Lessons learned from using a computer-based proficiency test of english for academic purposes. *Journal of Information Technology Education*, 2(1):185–201.
- Aquino, V. T. (2001). Avaliação automática de exames de proficiência em inglês. Mestrado, ICMC-USP, Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo, São Carlos, Brasil.
- Baker, F. (2001). *The Basics of Item Response*. ERIC Clearinghouse on Assesment and Evaluation, University of Maryland, College Park, MD, second edition.
- Bloom, B. S. and Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives, Handbook I: Cognitive Domain*. Addison Wesley Publishing Company.
- Bruno, J. E. (1986). Assessing the knowledge base of students: An information theoretic approach to testing. *Journal of Measurement and Evaluation in Couseling and Development*, 19:16–30.
- Conejo, R., Millán, E., Cruz, J. L. P., and Trella, M. (2001). Modelado del alumno: um enfoque bayesiano. *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial*, (12):50–58.

-
- Haydt, R. C. (1988). *Avaliação do Processo Ensino-Aprendizagem*. Editora Ática, S.A., São Paulo, Brasil.
- Huang, S. X. (1996). A content-balanced adaptive testing. in computer aided learning and instruction in science and engineering. *CALISCE-Computer Aided Learning and Instruction in Science and Engineering*, (3):29–31.
- Hutchinson, T. and Waters, A. (1987). *English for Specific Purposes*. Cambridge University Press.
- Klinger, A. (1997). Experimental validation of learning accomplishment. In *Presented at the ASEE/IEEE Frontiers in Education Conference July 11, 1997*. Retrieved September, 3, 2005.
- Lord, F. M. (1980). *Application of Item Response Theory to Practical Testing Problems*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, EUA, first edition.
- McKenna, C. and Bull, J. (1999). Design effective objective test questions: an introductory workshop. *Proceedings of the Conference at Loughborough University, Flexible Learning*, (Third):253–257.
- Miller, A., Imrie, B. W., and Cox, K. (1998). *Student Assessment in Higher Education, A Handbook for Assessing Performance*. Kogan Page, 120 Pentonville Road, London, UK, first edition.
- Monzón, A. J. (2008). Construção de banco de questões para exames de proficiência em inglês para programas de pós-graduação. Master's thesis, PPGL-UFSCar, Programa de Pós-Graduação em Linguística, Universidade Federal de São Carlos, São Carlos, Brasil.
- Olea, J., Ponsoda, V., and Prieto, G. (1999). *Tests Informatizados Fundamentos y Aplicaciones*. Ediciones Pirámide.
- Oliveira, L. H. M. (2002). Testes adaptativos sensíveis ao conteúdo do banco de itens: uma aplicação em exames de proficiência em inglês para programas de pós-graduação. Mestrado, ICMC-USP, Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo, São Carlos, Brasil.
- Pasquali, L. and Primi, R. (2003). Fundamentos da teoria de resposta ao item - tri. *Avaliação Psicológica*, 2:99–110.
- Piton-Gonçalves, J. (2004). A integração de testes adaptativos informatizados e ambientes computacionais de tarefas para o aprendizado do inglês instrumental. Master's thesis, ICMC-USP, Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo, São Carlos, Brasil.
- Piton-Gonçalves, J., Aluisio, S. M., Mendonca, L. H., and Novaes, O. O. (2004). A learning environment for english for academic purposes based on adaptive tests and task-based systems. *Lecture Notes in Computer Science*. Publisher: Springer-Verlag Gmb., 7:1–11. ISBN: 3-540-22948-5.
- Shuford, E. H. and Brown, T. (1974). Rationale of computer-administred admissible probability measurement. *Defense Advanced Research Projects Agency*, R-1371-ARPA.
- Weiss, D. J. and Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Education Measurement*, 21:361–375.