

Criação e aplicação de Testes Adaptativos Informatizados: um estudo de caso.

Leandro Henrique Mendonça de Oliveira¹, Sandra Maria Aluísio^{1,2}, Jean Pítton Gonçalves²

1 - Núcleo Interinstitucional de Linguística Computacional (NILC), CP 668, 13560 -970 São Carlos, SP, Brasil

2 - ICMC-Universidade de São Paulo, CP 668, 13560-970 São Carlos, SP, Brasil

leandroh@nilc.icmc.usp.br, sandra@icmc.usp.br, jpiton@icmc.usp.br

Abstract: *The main objective of the diagnostic evaluation is to measure the students' knowledge and classify them in the teaching programs. Based on Item Response Theory (IRT) the computer adaptive testing conjunctly with the content-balanced item bank is one of the most important implementation approaches in the computer assessment field. This paper presents the processes of creation and calibration of content-balanced item bank related to the English Proficiency Test (EPT) of the Instituto de Ciências Matemáticas e de Computação (ICMC) of the Universidade de São Paulo (USP). Specifically, we describe all activities related to the processes of development, design and calibration of the item bank.*

Resumo: *O principal objetivo da avaliação diagnóstica é medir o conhecimento do aluno e posicioná-lo adequadamente nos programas de ensino. Baseado nos modelos de resposta da Teoria de Resposta de Itens (TRI), o teste adaptativo informatizado em conjunto com um banco de itens sensível ao conteúdo, é uma das implementações mais elegantes no campo da avaliação informatizada. Este trabalho apresenta o processo de criação e calibração de um banco de itens sensível ao conteúdo referente ao Exame de Proficiência em Inglês (EPI) para os alunos da pós-graduação do Instituto de Ciências Matemáticas e de Computação (ICMC) da Universidade de São Paulo (USP). Especificamente, são descritas todas as atividades envolvidas no processo de desenvolvimento, estruturação e calibração do banco de itens.*

1. Introdução

Em linhas gerais a avaliação está inserida no contexto da educação assumindo importantes papéis que servem para corroborar o aprendizado e determinar o potencial de um indivíduo. Considerando estes aspectos é possível delinear diversas maneiras de avaliar um estudante, dentre elas a avaliação diagnóstica, cujo principal objetivo é posicionar adequadamente os estudantes nos programas de ensino que serão disponibilizados. Por essa razão os testes inerentes a esta avaliação são empregados nos estágios iniciais do processo de ensino, preferencialmente antes do início dos estudos, de maneira que as habilidades ou as deficiências de cada estudante possam ser identificadas.

Atualmente os estudantes que ingressam no programa de pós-graduação do ICMC-USP¹ são avaliados com relação à proficiência em inglês por meio de um teste que avalia a estrutura e convenções da língua inglesa para o gênero de artigos científicos, o que é essencial para o pesquisador desempenhar melhor e mais rápido as tarefas de leitura e escrita relacionadas à pesquisa [Swales, 1998]. O Exame de Proficiência em Inglês (EPI), implantado no formato de lápis e papel no segundo semestre de 2000, no formato computadorizado no primeiro semestre de 2001 [Aquino, 2001] e no computadorizado adaptativo em 2003 [Gonçalves e Aluísio, 2004], possui dois propósitos: 1) avaliar as habilidades de estruturar, compreender e reconhecer as relações de idéias de um texto além de conhecer as estratégias de escrita de um artigo científico e 2) tentar sanar os problemas de custo, tempo, falta de padronização, subjetividade e ausência de adequação nos exames tradicionais de proficiência.

O EPI é composto de quatro módulos, conforme descrito na Tabela 1, contemplando o conteúdo relacionado com as convenções da língua inglesa para o gênero de textos científicos.

Tabela 1 – Identificação dos módulos do EPI

Módulos do EPI
M1 – Convenções da Língua Inglesa para Textos Científicos
M2 – Estruturação de Textos Científicos
M3 – Compreensão de Texto
M4 – Estratégias de Escrita

Uma das implementações mais elegantes no campo da avaliação informatizada são os Testes Adaptativos Informatizados (TAIs). Estes testes empregam características específicas, descritas nos modelos da **Teoria de Resposta de Itens (TRI)** [Baker, 1998, Hambleton e Swaminathan, 1995, Rudner, 1998], como: o nível de dificuldade, o fator de adivinhação (chute) e a discriminação de cada uma das questões do teste, bem como a habilidade individual do estudante para estimar seu nível de rendimento [Olea et.al. 1973].

A idéia fundamental dos testes adaptativos é ajustar os itens de um teste ao nível de habilidade individual de cada participante de uma avaliação. Em outras palavras, é apresentada uma sequência de questões que mais se adaptam ao conhecimento de cada estudante, de maneira que um teste não será o mesmo para todos os indivíduos, possibilitando uma medida mais fiel a respeito da competência de cada um, bem como a economia de tempo de aplicação dos testes.

Os TAIs mais recentes utilizam um *banco de itens*, cujos itens são armazenados de forma estruturada, facilitando a busca e recuperação durante um teste. Um *banco de itens* pode ser planejado e organizado para cobrir diferentes conteúdos de um mesmo domínio de conhecimento, ou seja, podem ser criados grupos de itens para serem aplicados em diferentes momentos, medindo diferentes habilidades, em uma ou mais avaliações [Kingsbury e Zara, 1989, Olea et.al. 1973].

No entanto para isso ser possível, é necessário realizar uma “calibração” do banco de itens ajustando-o a um modelo de resposta (pertencente à TRI). É preciso escolher e avaliar quais os modelos de resposta mais indicados para que os objetivos da avaliação sejam atingidos. Assim, numa perspectiva otimista, os testes adaptativos são uma boa opção de avaliação,

¹ Instituto de Ciências Matemáticas e de Computação <http://www.icmc.sc.usp.br>

principalmente quando estes são apoiados em bancos de itens bem elaborados e calibrados. Entretanto, segundo Huang [1996] há dois obstáculos que fazem o emprego dos TAIs ser difícil: 1) a adoção de um estudo empírico de larga escala para uma boa calibração dos itens do banco e, 2) a geração de avaliações multidimensionais que atendem a diferentes conteúdos de uma mesma área de conhecimento.

Este artigo tem como principal objetivo descrever o processo de criação e desenvolvimento de um banco de itens calibrado e robusto, que forneça suporte para a implementação de uma avaliação diagnóstica sob a abordagem adaptativa que contemple os diversos módulos do EPI do ICMC-USP. O trabalho de [Gonçalves, et. al. 2004] que utilizou o método de calibração empregado no algoritmo CBAT-2, baseado em [Huang, 1996] também será discutido.

A seção 2 deste artigo mostra as principais características de um banco de itens, apresentando todo o processo de criação e calibração do banco de itens, variando desde sua organização até o resultado da estimação dos parâmetros, descrevendo as tarefas de coleta, análise e estruturação dos itens. Na seção 3 são apresentadas as principais características do **SisBI** (Sistema de Gerenciamento do Banco de Itens), ficando a seção 4 responsável pela descrição da modelagem **TAEPI**. Na seção 5 são discutidos os resultados obtidos e as conclusões deste artigo.

2. Criação e Calibração do Banco de Itens

A grande capacidade de processamento dos computadores atuais os tornaram em ferramentas extremamente úteis no campo da avaliação informatizada. Não só para o processamento dos resultados dos testes (pontuação, análise e elaboração de relatórios), mas para a construção, administração e armazenamento de itens.

Podemos definir um banco de itens com sendo um volumoso conjunto de itens (questões de teste) pertencentes ao mesmo domínio de conhecimento, que armazenados de forma estruturada, facilitam a sua recuperação em um determinado momento. Porém, esse aspecto do banco de itens exige que os itens armazenados se refiram a apenas um único conteúdo do domínio de conhecimento que o banco implementa. Entretanto, em uma dada área de conhecimento podem existir diversos conteúdos que necessitam ser avaliados em uma avaliação, pois o professor pode querer avaliar diferentes habilidades que dependem de conteúdos variados.

Dessa maneira, para atender os diversos conteúdos de uma determinada área do conhecimento, seria necessário criar vários bancos de itens, um para cada conteúdo da avaliação proposta. No entanto, existe uma técnica aplicada na construção do banco de itens que é identificada como *Balanceamento de Conteúdo*² [Olea et.al. 1973, Kingsbury e Zara, 1989, Wainer e Kiely, 1987]. Essa técnica permite a divisão do banco de itens em várias seções, cada uma delas representando um conteúdo específico que se deseja avaliar. O uso dessa técnica evita a “quebra” do banco de itens em vários pequenos bancos e permite que um único exame avalie várias habilidades dos estudantes sobre o mesmo domínio de conhecimento.

² Content-Balanced em inglês.

Outra técnica que pode ser implementada na construção do banco de itens é o uso de *Testlets*. Segundo Wainer e Kiely [1987] um *Testlet* é um grupo de itens relacionado a um determinado conteúdo que é desenvolvido como uma “unidade de teste” contendo um número de itens predeterminado, e que pode ser fornecido a um aluno durante uma avaliação.

A construção híbrida de um banco de itens que implementa o *Balanceamento de Conteúdo* e os *Testlets* pode ser muito proveitosa, já que é possível garantir que um determinado exame avalie várias habilidades. Tal característica é que permite a implementação de testes adaptativos sensíveis ao conteúdo.

2.1. Organização do Conteúdo do Banco de Itens

O conteúdo do banco de itens para a avaliação adaptativa do EPI foi organizado de acordo com a distribuição dos módulos da Tabela 1. Dessa forma, consideramos que cada módulo do exame corresponde a uma “habilidade” que será avaliada durante um teste. Por sua vez cada módulo foi dividido em partes e as questões (itens) foram distribuídas entre estas partes de maneira que tal estratégia garantiu o *balanceamento de conteúdo*, além da existência de *Testlets* cujos grupos de questões foram organizados em algumas das partes presentes nos módulos.

A organização do conteúdo do banco de itens de acordo com cada módulo e suas respectivas partes é apresentada na Figura 1. Observe que cada módulo representa uma divisão do conteúdo e suas respectivas partes podem ser vistas como especializações deste conteúdo tornando-o flexível no momento de uma avaliação.

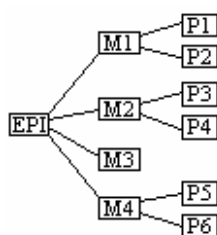


Figura 1 – Organização do conteúdo (módulos e partes) do banco de itens

2.2. Coleta dos Dados de Análise

A partir do primeiro semestre de 2001 o EPI passou a ser realizado de forma computadorizada³. Durante todo o ano de 2001 foram aplicados vários exames aos alunos da pós-graduação do ICMC, sendo que alguns deles tinham caráter de treinamento e outros de caráter definitivo e formal, cujos resultados eram considerados oficiais pelo programa de pós-graduação deste instituto. Portanto, como a intenção era construir e calibrar um banco de itens relacionados ao EPI, os resultados da aplicação destes exames foram utilizados, pois para o processo de calibração dos itens era preciso um número de ocorrências (alunos que fizeram um exame) considerável.

Ao todo foram coletados oito exames (provas), cada um com um número diferente de alunos que os realizaram. Cada prova pertence a uma dada área de pesquisa que é contemplada pela pós-graduação deste instituto.

³ <http://www.nilc.icmc.sc.usp.br/staff/capteap/>

2.3. Organização das Provas do EPI

Dado que o conteúdo do banco de itens é dividido em módulos (obedecendo a hierarquia da Figura 1) a primeira tarefa desempenhada tendo a posse das provas, foi identificar quais módulos e partes eram cobertas em cada uma delas. Tal tarefa produziu a Tabela 2, que mostra os módulos e partes cobertas por cada prova. Observe que os mesmos estão dispostos na mesma ordem que uma prova que EPI utiliza, e que de certa forma estão uniformemente distribuídos, cobrindo todos os módulos existentes. Repare ainda na Tabela 2 que um código M4-5 identifica o módulo 4 na parte 5.

Tabela 2 – Distribuição dos módulos e partes em cada prova

PROVA 01						PROVA 05					
M2-3	M4-5	M3	M1-2			M2-3	M2-4	M4-5	M3	M1-1	
PROVA 02						PROVA 06					
M2-3	M2-4	M4-5	M3	M1-1	M1-2	M2-3	M2-4	M4-5	M3	M1-1	
PROVA 03						PROVA 07					
M2-3	M2-4	M4-5	M4-6	M3	M1-2	M2-3	M2-4	M4-5	M3	M1-1	
PROVA 04						PROVA 08					
M2-3	M2-4	M4-5	M4-6	M3	M1-1	M1-2	M2-3	M2-4	M4-6	M3	M1-2

2.4. Análise e Seleção das Questões Válidas

Terminada a identificação dos módulos e partes contidas nas provas, foi realizada uma análise dos itens com a finalidade de identificar questões semelhantes entre todas as provas para separá-las em grupos relacionados com cada parte. Tal atividade permitiu a contabilização das questões, a formação dos conjuntos de estimação (formado por todas as questões pertencentes a uma parte) e o conhecimento dos itens que fariam parte do processo de estimação de parâmetros (calibração).

A tarefa de identificação dos itens semelhantes foi realizada através da observação cuidadosa de todas as questões das provas, sendo verificado parte por parte e questão por questão individualmente. O total geral de questões foi de 190 e o número de questões semelhantes foi 68, ou seja, o total de questões autênticas que deverão participar do processo de calibração foi de 122.

2.5. Elaboração dos Conjuntos de Dados para Estimação de Parâmetros

Encerrada a identificação e contabilização das questões, foi necessário elaborar vários conjuntos de questões que posteriormente serviriam para o processo de calibração. A tarefa de calibração de um banco de itens que implementa o balanceamento de conteúdo é realizada separadamente, sobre cada conteúdo que será incluído no banco. Assim, para o EPI, as questões referentes a cada módulo e parte foram estimadas de forma isolada, como se cada módulo fosse um banco de itens diferente.

Em razão da sensibilidade de conteúdo do banco de itens, houve a necessidade de uma organização das questões das provas em determinados grupos, que foram aqui chamados de *Conjuntos de Análise*, de maneira que as questões semelhantes foram agrupadas em um único conjunto, aproveitando as ocorrências de resposta (resposta dos alunos) de todas as provas a fim de melhorar o resultado da calibração. Assim, um *conjunto de análise* é um grupo de questões referentes a um mesmo módulo e parte contendo todas as ocorrências de resposta, sendo estas pertencentes a uma ou mais provas. Ao final desta fase foram definidos 31 *conjuntos de análise*,

alguns deles contendo um único grupo de questões referentes a uma única prova outros possuindo grupos de até quatro provas diferentes.

Generalizando, cada *conjunto de análise* pode ser entendido como sendo “grupos” de itens que serão calibrados automaticamente⁴, respeitando o balanceamento de conteúdo definido na estrutura do banco de itens.

2.6. O Processo de Calibração

A elaboração dos conjuntos de análise é a fase que antecede a calibração dos itens do banco. A estimação de parâmetros consiste na submissão dos itens a um método de estimação regido por um modelo matemático, geralmente executado por um programa de computador.

Para a estimação dos parâmetros das questões do EPI foi aplicado o método da **Estimativa de Máxima Verossimilhança** (EMV) [Baker, 1998, Meyer, 1973] sob o modelo logístico de dois e três parâmetros da TRI ($P_i(q)$) e pelo emprego do *Algoritmo EM (Expectation-Maximization)* [Bilmes, 1998, Woodruff e Hanson, 1997, Hanson, 1996] todos implementados pelo programa XCALIBRE [Xcalibre Manual, 1997].

Todos os conjuntos de análise foram formatados segundo o modelo de arquivo de entrada do XCALIBRE e submetidos ao processo de estimação. Cada conjunto de análise foi sujeito à estimação dos modelos de 2 e 3 parâmetros (2P e 3P). Após a estimação, foi realizada uma análise dos resultados deste processo (por meio dos relatórios finais do XCALIBRE) com intuito de selecionar qual o modelo melhor se ajustava às questões do EPI.

Entre os 31 conjuntos de análise elaborados e submetidos ao processo de estimação, apenas 3 não foram estimados por nenhum modelo (2P e 3P) por conterem poucas ocorrências. Dos 28 conjuntos restantes, 13 foram estimados pelo modelo de 2P e 15 pelos modelos de 2P e 3P. Decorrente desses conjuntos, as 122 questões que foram distribuídas nos conjuntos de análise e sujeitas à estimação, 113 foram estimadas, sendo que 42 para o modelo 2P e 71 para o modelo 2P e 3P. As 9 questões restantes pertenciam aos conjuntos (3) não estimados citados anteriormente. A análise dos relatórios finais foi realizada em apenas 15 conjuntos, ou seja, somente nas 71 questões que foram estimadas para os modelos 2P e 3P.

2.7. Resultados da Estimação dos Parâmetros

Baseado do relatório final do programa XCALIBRE a decisão de ajustar um determina do item a um modelo foi efetuada comparando os seguintes valores produzidos durante a fase de estimação: 1) o erro padrão associado a cada estimativa e 2) o valor residual padrão de cada item. Um determinado item, estimado para os modelo 2P e 3P, é escolhido quando estes valores de um dos modelos forem menores ou próximos de zero.

Uma pergunta que pode surgir é: *Porque alguns conjuntos de análise foram estimados para os modelos 2P e 3P e outros só para o modelo 2P?* Tal ocorrência pode ser entendida se pensarmos no comportamento de cada questão inserida no contexto da avaliação. Como a estimação se dá a partir da análise das respostas fornecidas às questões em uma situação real de teste, o que pode ser dito é que o volume massivo destes dados carrega o comportamento empírico das questões frente às suas aplicações em uma avaliação, devido a vários fatores como o ambiente de aplicação da avaliação, o nível de comprometimento dos alunos e seus objetivos.

⁴ Estimação utilizando o programa XCALIBRE (www.assess.com)

Tais fatores são diretamente refletidos nas respostas dos alunos, que é a base para estimação e cujo resultado reflete o comportamento dos itens. Portanto, uma nova estimação no futuro pode ser diferente, já que com o tempo as situações de aplicação de uma avaliação se alteram e seja refletida nos parâmetros estimados.

Apesar de 15 conjuntos de análise terem sido estimados para os modelos 2P e 3P, apenas 3 itens (destes conjuntos) ajustaram-se ao modelo 3P. O restante se ajustou ao modelo 2P. Finalmente, dos 122 itens que participaram do processo de estimação, 17 foram excluídos, ou por terem poucas ocorrências ou por terem sido excluídos durante a estimação; 3 itens foram estimados para o modelo de 3P e 100 itens para o modelo de 2P. Dessa maneira, temos um total de 103 questões calibradas, o que resulta no aproveitamento de 84,43%.

Finalizando todas essas tarefas temos como resultado um banco de itens referentes ao Exame de Proficiência em Inglês do ICMC, sensível ao conteúdo e calibrado para o modelo da TRI de 2 parâmetros. Nesse caso específico, como a maioria dos itens se ajustou ao modelo 2P, todos os itens foram armazenados no banco sob este modelo.

2.8. Uma alternativa para a calibração do banco de itens

Uma alternativa menos custosa e mais viável para a calibração do banco de itens e conseqüentemente a aplicação dos testes adaptativos foi empregado no trabalho de Gonçalves e Aluisio [2004]. Neste trabalho a calibração dos itens foi realizada empregando o algoritmo descrito em Huang [1996], chamado de Content Balanced Adaptive Testing (CBAT-2), que é a base de um teste “auto adaptativo” que calibra os itens do banco durante o acontecimento de um teste. A idéia deste tipo de calibração é interessante pois alivia a carga existente no processo de pré-calibração dos itens (exigente na aplicação dos testes adaptativos), além de considerar a performance dos estudantes. Entretanto, este algoritmo calibra somente o parâmetro b (dificuldade) baseando-se nas respostas dos alunos, após a definição do seu valor inicial por um especialista. Os outros parâmetros, c (adivinhação) e a (discriminação) assumem valores fixos, não sendo, portanto, calibrados.

Este método foi implementado como um módulo do sistema CALEAP -Web [Gonçalves e Aluisio, 2004]. Este sistema foi baseado no TAEPI, cujas funcionalidades serão descritas na seção 4, e proporcionou a aplicação da avaliação adaptativa no ambiente do ICMC-USP.

3. Sistema de Gerenciamento do Banco de Itens (SisBI)

A criação do banco de itens, descrita na seção anterior, seria pífia se não possuíssemos uma ferramenta para a manipulação do mesmo. Com o intuito de aumentar a organização do banco de itens foi proposto o **Sistema de Gerenciamento do Banco de Itens (SisBI)**, cujo principal objetivo foi facilitar a manutenção do banco, auxiliando o professor na tarefa de gerenciar os itens de forma simples e otimizada. A manipulação de questões do EPI voltada para a avaliação adaptativa é baseada no ambiente Web, tornando-se uma vantagem, já que o sistema depende somente de um *Browser* para ser executado. Pelo fato do **SisBI** ser tipicamente um sistema acadêmico, voltado especialmente para professores, seu desenvolvimento se deu com o uso de ferramentas de domínio público, favorecendo assim sua posterior atualização e crescimento conforme seja necessário. A modelagem do sistema foi desenvolvida usando a **UML** (*Unified Modeling Language*) e a implementação foi utilizada a linguagem **PHP** (*Personal Home Page*) em conjunto com o banco de dados **MySQL** sobre o *servidor internet Apache*.

O **SisBI** é formado por um único ambiente composto de funcionalidades voltadas para a manutenção de um banco de itens. O único usuário que interage com o sistema é o professor que desempenha as tarefas de inclusão, exclusão, alteração e visualização de questões, bem como a edição de textos e partes (divisão de conteúdo) que podem conter em um exame. Outra funcionalidade desempenhada pelo sistema é a montagem dos gráficos relacionados com a TRI (*Função de Informação de um Item* e *Função de Resposta de Item* [Baker, 1998]) de cada item pertencente ao banco. Tal funcionalidade permite que o professor verifique o comportamento do item, dado os gráficos das funções de resposta e de informação.

4. Modelagem do TAEPI

A criação do banco de itens e o desenvolvimento do **SisBI** estimularam o trabalho de modelagem de um sistema adaptativo de avaliação diagnóstica da proficiência em inglês do programa de mestrado do ICMC-USP, batizado de TAEPI - *Testes Adaptativos para o Exame de Proficiência em Inglês*. O principal objetivo dessa modelagem é descrever os ambientes e funcionalidades referentes a um teste adaptativo voltado para este exame de proficiência em inglês, permitindo a implementação e aplicação do mesmo no cenário do ICMC-USP. Assim como foi feito na modelagem do **SisBI**, o mapeamento das classes, relacionamentos de tabelas e funções do TAEPI também foram desenvolvidos em UML (*Unified Modeling Language*), cujo enfoque foi à definição de diagramas necessários para a concepção do sistema. Por meio dos diagramas da UML foi possível representar o sistema sob diversas perspectivas de visualização, facilitando o entendimento do seu comportamento.

As funcionalidades específicas oferecidas pelo TAEPI variam de acordo com os três tipos de usuários (Administrador, Professores ou Alunos) que podem utilizá-lo. Deste modo, pode ser entendido que o sistema disponibiliza diferentes *ambientes de operação*, dependendo somente do usuário que vai utilizar.

Dessa maneira, o TAEPI é composto por três ambientes distintos de operação, sendo estes, um ambiente para o *Administrador*, uma para os *Professores* e outro para os *Alunos*. Cada ambiente é formado por tarefas (funcionalidades) que são de responsabilidade de cada usuário pertencente àquele ambiente, as quais seguem:

- **Administrador:** responsável pelo cadastro de professores e gerenciamento do sistema ;
- **Professores:** responsável principalmente pela criação e especificação de exames adaptativos que serão realizados pelos alunos, além da possibilidade de consulta dos mesmos;
- **Alunos:** são responsáveis pelo seu próprio cadastro e sua principal tarefa é realizar um exame adaptativo especificado por um professor, além de poder consultar seus resultados e gabaritos de exames já realizados.

A seguir serão descritas as funcionalidades de cada ambiente.

4.1. O Ambiente do Administrador

O administrador, no contexto do sistema TAEPI, possui duas funções: 1) o cadastramento do professor e 2) o gerenciamento do sistema.

O cadastramento de professores possibilita que professores interessados em utilizar o TAEPI criem e especifiquem exames adaptativos.

4.2. O Ambiente do Professor

Este ambiente contém todas as tarefas que um determinado professor (previamente cadastrado) pode realizar dentro do **TAEPI**. São elas:

- **Efetuar Login:** corresponde à permissão ou não de acesso do professor ao sistema. Acontece quando um professor abre a tela de acesso do sistema e fornece um *username* e *Senha* que são verificados;
- **Criar Exame:** possibilita um determinado professor criar e definir um exame adaptativo que será disponibilizado aos alunos. Dentre as informações que o professor deve fornecer para a criação de um exame, estão as partes (módulo e parte) que irão compor o exame;
- **Especificar Critérios de Aplicação do Exame:** após a criação de um exame, o professor pode especificar seus critérios de acordo com as partes inseridas no mesmo. Tal especificação é bastante flexível e permite a definição de valores máximos e mínimos de habilidade do exame, a quantidade de questões que se deseja administrar e ainda a quantidade de *testlets* (caso alguns dos módulos inseridos contenha) requeridos, bem como os valores máximo e mínimo da habilidade de cada *testlet*. Para especificar os critérios de aplicação, o professor deve fornecer um número de identificação do exame previamente cadastrado .
- **Consultar Resultado Global:** a consulta do resultado global acontece quando o professor fornece a identificação do(s) aluno(s) ou um determinado exame, permitindo que um professor verifique os resultados de um determinado exame que já foi realizado. O **TAEPI** oferece as opções de verificar o resultado individual ou coletivo dos alunos. Caso esta última opção seja selecionada, o professor também tem acesso à média do valor da habilidade e a pontuação alcançada em cada parte pelo aluno;
- **Consultar Resultado Detalhado:** por meio desta função o professor pode verificar o resultado detalhado de um determinado aluno em um exame já realizado. O resultado detalhado contém informações de todas as questões fornecidas ao aluno durante um exame adaptativo, juntamente com o módulo e parte que pertence, a resposta fornecida, os cálculos parciais dos valores de habilidade e pontuação obtida. O professor deve fornecer as identificações do exame e do aluno para obter estes resultados.

4.3. O Ambiente do Aluno

O ambiente do aluno possui todas as funções para resolução de exames adaptativos previamente cadastrados pelo professor. Além da possibilidade de realizar um exame (função *Realizar_Exame*) os alunos também podem se cadastrar, consultar resultados de exames e consultar o gabarito de um determinado exame realizado anteriormente. Abaixo são descritas todas as atividades que os alunos podem desempenhar no TAEPI.

- **Cadastrar Aluno:** possibilita o cadastramento de alunos que irão realizar um determinado exame de proficiência em inglês. O cadastro é realizado pelo próprio aluno quando este acessa pela primeira vez o sistema e escolhe a opção *cadastrar*, fornecendo algumas informações como nome, *username*, senha entre outros. O cadastro de aluno permite o mesmo realizar qualquer exame que esteja disponível no sistema;

• **Efetuar Login:** permite o acesso do aluno ao sistema de maneira que este possa desempenhar as tarefas do ambiente. Semelhante ao acesso do professor, consiste no fornecimento do *username* e senha que são verificados;

• **Realizar Exame:** é a principal funcionalidade que este ambiente proporciona, pois permite que um aluno previamente cadastrado realize um exame especificado pelo professor. Resumidamente, a realização de um exame é a execução pura da avaliação adaptativa, e nela contém todos os procedimentos de seleção de itens, estimação de habilidade e verificação dos critérios de parada que permitem a tarefa de adaptação exercida pelo exame;

• **Consultar Resultado:** permite que um determinado aluno verifique os resultados de um exame já realizado por ele. O aluno deve fornecer a identificação do exame realizado e sua senha de acesso, de maneira que após a autenticação, o sistema exibe as seguintes informações: número do exame, módulos e partes avaliadas, quantidade de questões feitas, valor da habilidade e pontuação finais;

• **Consultar Gabarito:** depois que um determinado aluno realizou um exame, é permitido que ele verifique o gabarito do mesmo, de maneira que ele possa conhecer os erros e acertos cometidos. Assim, o aluno fornece o número do exame realizado e recebe como resposta as informações de identificação da parte, a questão fornecida, a resposta dada e a resposta correta. Entretanto, tal função deve ser muito planejada antes de ser disponibilizada, pois ela expõe os itens do banco, o que pode ser muito prejudicial para uma avaliação adaptativa formal.

4.4. A Importância do TAEPI

A modelagem do protótipo do sistema de avaliação adaptativa (TAEPI) procurou explorar todas as vantagens e potencialidades proporcionadas pelo balanceamento de conteúdo do banco de itens. É possível criar testes adaptativos contendo um único ou vários módulos pertencentes ao banco de itens e, conseqüentemente, especificar seus critérios de parada de maneira individualizada. Portanto, com intuito de se beneficiar da sensibilidade de conteúdo do banco de itens, a modelagem do sistema TAEPI prevê a criação de testes adaptativos multidimensionais (como é o caso do CALEAP-Web), por meio da especificação de diversos módulos e partes que podem estar presentes em uma avaliação, além da definição de variados critérios de parada para cada conteúdo inserido em um exame adaptativo. Dessa maneira, podemos dizer que a liberdade de especificar exames contendo um ou vários “conteúdos” (representados pelos módulos e partes) e em cada um destes determinar seus critérios de parada, dá a possibilidade da criação de exames totalmente flexíveis e de acordo com a necessidade de uma avaliação. Logo, o que torna o sistema TAEPI interessante é autonomia de produzir exames maleáveis, condizentes com a realidade do ambiente de aplicação do mesmo.

É importante ressaltar que o TAEPI serviu como estímulo inicial para o desenvolvimento de sistemas de aplicação de testes adaptativos, servindo principalmente como “modelo” genérico de criação desses sistemas. Um exemplo disso foi a criação do CALEAP-Web que adveio das idéias dessa modelagem.

5. Conclusão

Geralmente, o processo de implantação da avaliação adaptativa no ambiente de pequenas instituições, seja empresarial ou acadêmica, torna-se difícil e demorado, pois o estudo empírico para o processo de calibração dos itens e a criação de testes sensíveis ao conteúdo são grandes obstáculos [Huang, 1996]. Entretanto, em centros de educação especializada, como o ETS (*Educational Testing Service*) tais obstáculos podem ser superados, já que nestes existe uma larga experiência vinda de anos de trabalhos dedicados às avaliações e grandes volumes de dados para realizar a calibração dos itens.

Para tentar superar estes obstáculos e implantar os testes adaptativos, a tarefa de estimação dos parâmetros dos itens no processo de calibração do banco deve seguir 4 passos (descritos em [Hambleton e Swaminathan, 1995]) dentre eles a análise de um conjunto de amostras de respostas de itens aplicados anteriormente. Contudo, a aplicação do método descrito em Huang [1996] e aplicado no ambiente do ICMC-USP no trabalho de Gonçalves [2004], também mostrou resultados satisfatórios, principalmente no contexto de pequenas instituições. Assim, a adoção destas metodologias é justificada pela tendência de crescimento do banco de itens, dado que no cenário do ICMC-USP novos EPIS acontecem a cada semestre, de maneira que novas questões serão calibradas e inseridas no banco.

Por sua vez, o ambiente CALEP-Web, baseado na modelagem do TAEPI, é o primeiro e mais importante passo em direção da implantação da avaliação adaptativa em pequenas instituições, pois integra a avaliação adaptativa com a calibração do banco de itens em um ambiente único de tarefas, servindo tanto para a avaliação dos estudantes como manutenção da avaliação adaptativa. No entanto, uma das maiores limitações do CALEP-Web é o tamanho do banco de itens, já que o EPI implementa o balanceamento de conteúdo, seu crescimento linear depende diretamente do crescimento de cada módulo e parte, pois é necessário que o módulo cresça de acordo com sua importância no contexto da avaliação.

Dessa maneira é fato que o principal componente de um sistema de avaliação adaptativa é o banco de itens. Considerado como elemento central do mesmo, podemos concluir que a vigência e implantação de um teste adaptativo ficam prejudicadas sem a existência de um banco de itens robusto e calibrado de acordo com as necessidades e objetivos da avaliação pretendida.

O processo de criação e calibração de um banco de itens sensível ao conteúdo (referente à proficiência em língua inglesa), bem como os aspectos envolvidos na estimação de parâmetros para os modelos de respostas da TRI é o principal resultado deste trabalho. O que se pôde aprender dessa experiência é que há a necessidade constante (das pessoas envolvidas) de *cuidado, planejamento, disciplina e organização*. *Cuidado* no que tange às decisões iniciais do projeto de construção do banco, correspondente àquelas que definem a estrutura e conteúdo do mesmo. *Disciplina* e *organização* são indispensáveis para a fase de calibração dos itens, pois é necessário ser coerente em relação à seleção do modelo de resposta. Por sua vez, é preciso *planejamento* para o desenvolvimento do banco, já que seu crescimento e manutenção são processos que exigem tempo e dedicação.

Referências

Aquino, V. T. (2001) “Avaliação automática de exames de proficiência em inglês”, Dissertação de Mestrado, Universidade de São Paulo – USP, São Carlos, 2001.

- Baker F. (1998) "Item Response Theory Parameter Estimation Techniques", MARCEL DEKKER, INC, 1998.
- Bilmes, J. A. (1998) "A gentle tutorial of the algorithm and its application to parameter estimation mixture and hidden markov models", Technical Report TC -97-021, International Computer Science Institute, 1998.
- Gonçalves, J. P. (2004) "A integração de Testes Adaptativos Informatizados e Ambientes Computacionais de Tarefas para o aprendizado do inglês instrumental", Dissertação de mestrado, ICMC-USP, São Carlos, Brasil, 2004.
- Gonçalves, J. P., Aluísio, S. M., Oliveira, L. H. M., e Oliveira Jr. O. N. (2004) "A Learning Environment for English for Academic Purposes based on Adaptive Tests and Task-based Systems" In: Proceedings of ITS' 2004, (no prelo)
- Hambleton, R. K. e Swaminathan, H. (1985) "Item Response Theory: Principles and Applications", Kluwer Nijhoff, 1985.
- Hanson, B. (1996) "Estimating parametric continuous latent distributions", Author's Note, 1996.
- Huang, S. X. (1996) "A content-balanced adaptive testing algorithm for computer-based training systems", In: ITS' 96 International Conference in ITS, Montral, 1996.
- Kingsbury, G. G. e Zara, A. R. (1989) "Procedures for selecting items for computerized adaptive tests", In: Applied measurement in education, 1989, pp. 16.
- Meyer, P. L. (1973) "Probabilidade: aplicações à estatística", Ao Livro Técnico S.A., 1973.
- Olea J., Ponsoda, V. e Prieto, G. (1999) "Tests Informatizados: Fundamentos y Aplicaciones", Ediciones Pirámide, 1999.
- Rudner, M. L. (1998) "An on-line, interactive, computer adaptive testing mini-tutorial", 1998.
- Swales, J. (1990) "Genre Analysis – English in academic and research settings", Cambridge University Press, 1990.
- Wainer, H. e Kiely, G. L. (1987) "Item clusters and computerized adaptive testing: a case of testlets", In: Journal of Educational Measurement, 1987, pp. 16.
- Woodruff, D. e Hanson, B. (1997) "Estimation for item response models using the EM algorithm for finite mixtures", In: Annual Meeting of the Psychometric Society, Tennessee, 1997.
- Xcalibre Manual (1997) "User's manual for the XCALIBRE marginal maximum-likelihood estimation", Assessment Systems Corporation (<http://www.assess.com>), 1997.