

ESTUDO COMPARATIVO DE ENTRE TRÊS APLICATIVOS UTILIZADOS PARA ANÁLISE DE ITENS DE PROVAS OBJETIVAS

Raimundo Helio Leite

UFC – rhleite@terra.com.br

Andréa Soares Rocha da Silva

UFC – andrea@elasanet.com.br

Introdução

No Brasil, a formação de pessoal para a área de magistério se faz nas Escolas Normais, nas Faculdades de Educação, nos Cursos de Licenciatura, e nos Cursos de Especialização, Mestrado e Doutorado. No caso das Escolas Normais e das Faculdades de Educação, os futuros professores cursam disciplinas e recebem informações sobre uma área específica que trata do problema de elaboração de provas e testes de rendimento escolar. Essa área é denominada de Medidas Educacionais. Fora disso, as informações sobre esse assunto são escassas, quando, não raro, inexistentes.

Os que militam na área educacional conhecem muito bem os dados sobre professores leigos, isto é, que não tem formação pedagógica adequada, que atuam no sistema educacional nos seus vários níveis. Deduz-se, facilmente, a quantas anda não só o processo ensino-aprendizagem, mas as formas de avaliá-lo. Essa atividade se processa, geralmente, através do efeito demonstração. Isto é, o futuro professor passa a desenvolver o mesmo método de elaboração de provas a que foi submetido quando estudante. É o método do aprender fazendo. As distorções são evidentes.

Apesar de essa situação ser generalizada, é possível se encontrar em nível de vestibular e em sala de aula, algumas boas exceções. Talvez premidas pela pressão da opinião pública, muitas Comissões Coordenadoras de Vestibular (Leite, 1974)



desenvolveram verdadeiros sistemas de avaliação de suas provas (Farias, 1974), utilizando-os para melhorarem seus exames e para criarem Bancos de Itens (Normando, 1974). Há casos, também, de professores que, individualmente, utilizam informações de medidas educacionais com o mesmo propósito.

O SAEL foi idealizado para funcionar como uma ferramenta de apoio ao desenvolvimento de instrumentos avaliativos de qualidade, sendo útil especialmente àqueles professores que não tiveram uma formação pedagógica adequada e, portanto, podem ser considerados leigos em relação a medidas educacionais.

Descrição Analítica dos Aplicativos

Esta seção descreve como especificidades de cada um dos aplicativos estudados, em função da teoria que lhe servem suporte. A discussão inicia-se com para de uma apresentação faz SAEL.

O Sistema de Avaliação Eletrônico (SAEL)

Antes de se fazer uma descrição do SAEL, uma primeira distinção deve ser feita entre três termos que são usados, não raro, sem a devida compreensão dos seus significados. Esses termos são **medir**, **testar** e **avaliar**.

Testar refere-se à atividade de aplicar testes. Esse conceito é muito útil quando se pretende padronizar um teste, isto é, estabelecer normas e padrões de rendimento do teste. **Medir** é associar números ou conceitos para representar atributos de objetos. **Avaliar** é comparar várias medidas sobre um mesmo atributo, com o objetivo de identificar semelhanças ou diferenças desse atributo. As duas últimas características são muito importantes para o estudo de variáveis associadas ao processo educacional. Por essa razão elas serão discutidas de forma mais acurada.

A **Medida Educacional**, para ser levada em conta de forma correta, deve satisfazer às exigências vistas anteriormente. Supondo que essas exigências sejam satisfeitas o próximo passo é tratar da forma como se realiza a medida em educação, isto é, tratar da elaboração de testes.

O SAEL utiliza a Teoria Clássica dos Testes (TCT), em que pese apresentar, Segundo Hambleton et al. (1991), as seguintes limitações teóricas:

- 1) Os parâmetros clássicos dos itens (dificuldade e discriminação) dependem diretamente da amostra de sujeitos utilizada para estabelecê-los (*group-dependent*).
- 2) A avaliação de aptidões dos testandos também depende do teste utilizado (*test-dependent*).
- 3) A definição do conceito de fidedignidade ou precisão na teoria clássica dos testes constitui também uma fonte de dificuldades.
- 4) Outro problema da teoria clássica dos testes consiste em que ela é orientada para o teste total e não para o teste individual.

De forma resumida, como afirma Fernandez (1990), "*En la teoria clásica, el resultado de la medicion de una variable depende del test utilizado, lo que plantea serios problemas para tratar de establecer la equivalencia entre las puntuaciones de los tests distintos que midam una mesma variable*". Essa restrição se faz presente muito freqüentemente no uso de medidas psicológicas. No que toca a medidas de rendimento escolar, especialmente em sala de aula, a situação é bem diferente. É sempre possível, adotando certos critérios e cuidados indicados pela teoria clássica, se elaborar testes de rendimento que cumpram sua finalidade pedagógica.



Indicadores da Qualidade de um Item

Além de possuir a precisão como indicador de ordem geral, um teste possui indicadores relacionados com seus componentes, isto é, as questões ou itens. Apresenta-se, a seguir, esses indicadores.

O **Índice de Dificuldade** representa o grau de acerto dos alunos que tentaram o item. Ele é dado por:

$$I_d = \frac{A_c}{N}$$

onde:
 A_c = número de alunos que acertaram a questão.
 N = número de alunos que tentaram a questão

Vê-se, facilmente, que $0 \leq I_d \leq 1$.

$I_d = 0$ quando nenhum dos alunos, dos que tentaram a questão, a acertam.

$I_d = 1$ quando todos os alunos que tentaram a questão, a acertam.

O **Índice de Poder Discriminante (IPD)** representa o nível de discriminação entre os alunos fracos e os alunos bons. O SAEL utiliza os 27% superiores e os 27% inferiores, adotando com isso os resultados obtidos por Kelley [1939] que mostrou que essa percentagem apresenta a melhor sensibilidade no cálculo desse indicador. Os melhores 27% dos alunos são denominados Massa Superior, e os 27% dos alunos mais fracos, Massa Inferior. Portanto,

$$IPD = I_d (\text{Massa Superior}) - I_d (\text{Massa Inferior})$$

O **Índice de Homogeneidade (IH)** representa a contribuição que o item (questão) dá para a nota total do aluno no teste. O uso adequado do IH possibilita se construir um teste, eliminando-se os itens que pouca ou nenhuma contribuição oferecem à nota que o aluno obtém. Vê-se que o IH é uma correlação entre o item e a nota total. Ele é calculado por:

$$t_{**} = \frac{\bar{X}_1 - \bar{X}_0}{s_1} \sqrt{\frac{n_1 n_0}{n(n-1)}}$$

Onde: X_1 = média dos que obtiveram resposta certa
 X_0 = média dos que obtiveram resposta errada
 n = número total de alunos que fizeram o teste
 n_1 = número de alunos com respostas certas
 n_0 = número de alunos com respostas erradas
 s_1 = desvio padrão das notas

O objetivo do SAEL, portanto, é colocar à disposição do professor as informações mínimas necessárias para elaborar, de forma adequada, seus testes de rendimento escolar, utilizando, para isto, elementos como: Tabela de Especificação, Análise de Item, Índice de Dificuldade e Índice de Homogeneidade, Índice do Poder Discriminante, Estatísticas Descritivas, além de um banco de dados e de uma ferramenta gráfica como componentes de apoio para elaboração de provas e de questões.

Análise e projeto do sistema

Para o desenvolvimento do SAEL utilizou-se a tecnologia orientada a objetos nas etapas de análise e de programação (implementação) do sistema. Uma ferramenta *Case* foi utilizada na etapa de análise, para auxiliar o processo de design e documentação do sistema, facilitar sua manutenção ou até mesmo futuras re-análises do sistema.

Os módulos do sistema foram desenvolvidos através da Linguagem de Programação Delphi. O Delphi foi escolhido como linguagem de programação por ser uma linguagem visual, orientada a objetos, que nos permite criar e adicionar ao seu repositório de objetos, os nossos próprios componentes (Pacheco & Teixeira, 1996). Um exemplo de tela de cadastro de questões do SAEL pode ser vista na figura 01.

Principais componentes do sistema

O SAEL está dividido em módulos funcionais, cuja execução é ilustrada na figura 02, e cuja descrição é feita logo em seguida.



Os principais módulos funcionais do SAEL são descritos a seguir:

- **Cadastros:** Alunos, Professores, Cursos, Disciplinas, Assuntos e Questões.
- **Elaboração de Provas:**

Consiste de três etapas:

- 1) **A elaboração da Tabela de Especificação.**
 - 2) **A inclusão de dados específicos sobre a prova** que está sendo elaborada, tais como: tempo de duração da prova, número de questões da prova, tipo da prova (objetiva, subjetiva ou mista), etc.
 - 3) **A seleção e/ou elaboração das questões**, onde o professor poderá selecionar uma questão, de acordo com os parâmetros definidos na etapa 1, ou elaborar uma questão “inérita”, utilizando, para isso, recursos de edição de textos e de editoração gráfica, que serão disponibilizados pelo SAEL.
- **Cadastro de Resultados para a Análise:** Esta função consiste de uma tela de entrada de dados, através da qual o professor fornecerá os resultados da aplicação da prova (para cada aluno, as questões acertadas e as questões erradas, além da nota obtida), para serem utilizados no procedimento de Análise Técnica da Prova.
 - **Mapa de Notas:** Esta função processa, exibe e imprime o mapa de notas dos alunos de uma determinada disciplina.
 - **Análise Técnica da Prova:** Esta é a principal função do SAEL. Aqui são aplicados os testes que avaliarão as provas a partir do desempenho dos alunos.
 - **Gráficos Estatísticos:** Esta função oferece uma descrição visual, através de gráficos, de alguns cálculos estatísticos feitos sobre as notas dos alunos.

- **Calculadora:** Função que exibe a calculadora do *Windows* para que o usuário possa fazer cálculos eventuais.
- **INFO:** Esta função, além de ser a função de ajuda do sistema ("*Help*"), explicando e exemplificando todos os cálculos e funções da análise técnica da prova, e exibir um Gráfico que mostra a seqüência de funcionamento e utilização do sistema, também oferece ao usuário um **minicurso de medidas educacionais**.

A função de meta-avaliação do SAEL se dá através do Relatório de Análise Técnica da Prova, o qual analisa cada item da prova identificando o seu índice de dificuldade, o seu índice de homogeneidade e o seu índice de poder discriminante. Através desses índices o professor verifica a qualidade das questões (itens) elaboradas e se estas correspondem ou não aos objetivos educacionais por ele definidos para os mesmos.

O Aplicativo MULTILOG

O MULTILOG, que realiza o cálculo de múltiplas funções logísticas, é um aplicativo projetado para proceder a análise de itens de testes de rendimento escolar e de testes psicológicos com base na de Teoria de Resposta de Item (TRI). Essa teoria calcula o índice de dificuldade de um item definido como o ponto numa escala de habilidade que define sua a probabilidade de acerto no item, o índice de discriminação definido como o ponto de flexão da curva característica do item e o acerto por acaso, entendido como a probabilidade de o aluno acertar o item sem saber o seu conteúdo. Vê-se, portanto, que as definições dos indicadores da qualidade de um item na TRI diferem das adotadas pela Teoria Clássica dos Testes.

O MULTILOG é usado em testes de múltipla escolha e trabalha com modelos logísticos (Samejina, 1969), o modelo para respostas classificadas (Bock, 1972) e o modelo para itens múltiplas escolhas (Thissen, 1986).



O Pacote Estatístico para Ciências Sociais (SPSS)

O SPSS é um dos mais populares pacotes dentre os aplicativos utilizados para se realizar análises quantitativas de dados em ciências sociais. Além de ter uma interface amigável, ele faz robustas análises. As rotinas deste programa permitem o cálculo dos indicadores usados pela TCT, com diferentes graus de dificuldade.

O Índice de Dificuldade (I_d) pode ser calculado diretamente via percentagem de acertos no item. Já o Índice de Homogeneidade requer bastante cálculo e conhecimento aprofundado desse aplicativo, pois o pesquisador terá que ordenar as notas, usando o recurso *Sort Cases* e, em seguida, selecionar o grupo de alunos que está nos 27% superiores e os que estão nos 27% inferiores com o emprego do recurso *Select Case*. Esse aplicativo calcula o Índice de Homogeneidade e a fidedignidade do teste, usando o *alfa* de Cronbach.

A utilização de *scripts* especialmente elaborados pelo pesquisador permite que o SPSS calcule os parâmetros dos modelos da TRI.

Outra vantagem do SPSS é que a definição prévia das variáveis permite a geração de tabelas que apresentam os dados em formato que pode ser diretamente usado para a geração do relatório da pesquisa. A Tabela a seguir mostra um resumo comparativo dos aplicativos discutidos.

Reflexões conclusivas

A discussão mostrou as virtudes e as fraquezas de três aplicativos que avaliam a qualidade de questões de provas e testes de rendimento escolar. Ficou claro que a Teoria de Resposta ao Item, por ter suporte teórico baseado em modelos probabilísticos, elimina os problemas decorrentes da aplicação em determinado grupo de aprendizes, tornando seus resultados mais consistentes e fidedignos. Em decorrência disso, a TRI

é adequada para utilização em situações de classificação e seleção de grandes grupos de pessoas como exames vestibulares e exames nacionais.

O SPSS é, tipicamente, uma ferramenta para se realizar cálculos estatísticos simples e complexos, função que desempenha com eficiência se o pesquisador souber as análises que pretende realizar.

O SAEL, por seu turno, reúne, num mesmo *software*, duas funções: a de treinamento de professores na técnica elaboração de testes, na medida em que reúne informações sobre elaboração de provas e permite que o professor realize análise da qualidade os itens; e a de criação de um Banco de itens de qualidade que podem ser reutilizados em outros testes. Além disso, através do SAEL, até mesmo algumas tarefas administrativas são facilitadas, uma vez que o aplicativo gera relatório com as notas para ser entregue no órgão responsável pelo registro da situação escolar dos alunos.

Por tudo isso, o SAEL parece ser uma alternativa válida para utilização em situação em sala de aula, e, se consideramos a sua função de treinamento de professores na elaboração de testes de rendimento escolar, pode ser especialmente útil para professores sem formação em medidas educacionais.

Referências Bibliográficas

BOCK, R. D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrica*, 37, 29-51, 1972.

FARIAS, O.T. Relatório Final – Seminário de Português para Análise da Prova Aplicada no 1o Vestibular/73 – Simpósio sobre Concurso Vestibular, Fevereiro de 1974.

FERNANDEZ, J.M. *Teoría de Respuesta a los Ítems*. Edições Pirâmide, S.A. Madrid, 1990.



HAMBLETON, R.K., SWAMINATHAN, H., ROGERS, H.J. *Fundamentals of Item Response Theory*. London: Sage, 1991.

KELLEY, T.L. *The selection of upper and lower groups for the validation of test items*. Journal of Educational Psychology, págs. 17-24, 1939.

LEITE, R.H. "A Precisão dos Exames Vestibulares na UFC" – Estudo Preliminar. Conselho de Reitores das Universidades Brasileiras Nacional – Simpósio sobre Concurso Vestibular, págs. 1-23, 1974.

NORMANDO, R.A. *Análise da Prova de Física do 2º Vestibular – 73 da UFC*. Simpósio Sobre Concurso Vestibular, págs. 1-35, 1974.

PACHECO, X.; TEIXEIRA, S. *Delphi 2 Developer's Guide*. Second Edition. Indiana: SAMS Publishing, 1996.

SAMEJINA, F. *Estimating of latent ability using a response pattern of graded scores*. (Psychoetric Monograph No 17), Iowa City, IA: Psychometric Society, 1969.

THISSEN, D. M. MULTLOG: Item analysis with multiple category response models (Version 5). Moorsville, In: *Scientific Software*, 1986.

FIGURAS E TABELAS

Figura 01 – Tela do Cadastro de Questões do SAEL

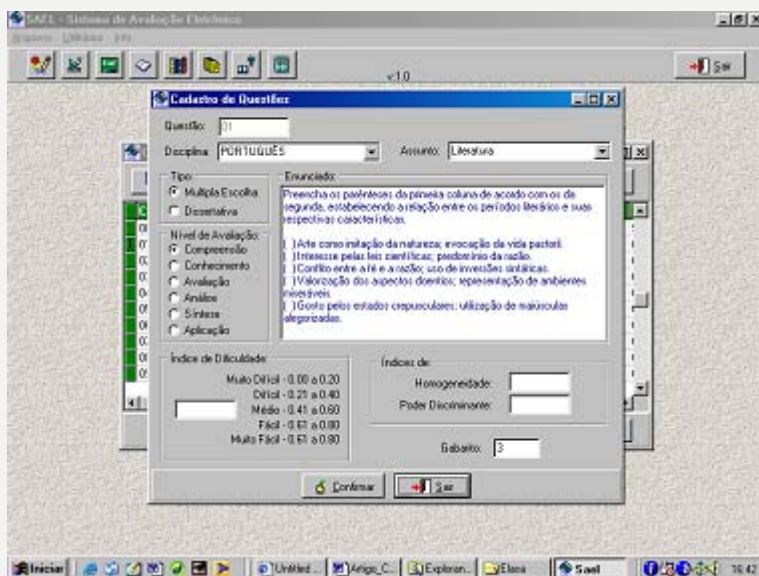


Figura 02 – Fluxo de Execução do SAEL

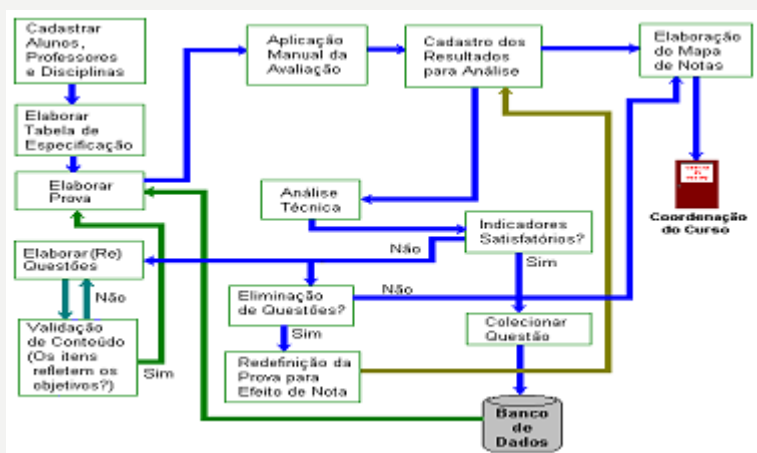


Tabela 01 – Resumo comparativo (x = alfa de Cronbach)

APLICATIVO	SAEL	MULTILOG	SPSS
Índice de Dificuldade	TCT	TRI	TRI/TCT
Índice de Discriminação	TCT	TRI	TRI/TCT
Índice de Homogeneidade	TCT	-	TCT
Acerto ao Acaso	-	TRI	-
Fidedignidade	x		x
Estatísticas Descritivas e Gráficos	Sim	Sim	Sim