Análise Métrica de um Instrumento de Medida da Aprendizagem Através da Teoria de Respostas ao Item (TRI)[1]

Wagner Bandeira Andriola[2]
José Anchieta Esmeraldo Barreto[3]

RESUMO

Esta pesquisa é resultado do processo de avaliação da aprendizagem dos alunos de escolas públicas do Estado do Ceará, realizado no final de 1996. O objetivo foi a estimação dos parâmetros métricos dos itens componentes do teste de português destinado aos alunos da 4ª série, através da Teoria de Respostas ao Item (TRI). Utilizando-se as respostas de 17576 alunos, cuja idade média foi 12,27 anos (DP=1,94), estimou-se a dificuldade e a discriminação dos 25 itens componentes do teste de português através do modelo logístico de dois parâmetros, determinando-se, também, as curvas características dos itens (CCI's) e do teste (CCT).

Wagner Bandeira Andriola
Mestre em Psicologia Social e do Trabalho pela UnB.
José Anchieta E. Barreto
PhD pela Universidade de Tulane - EUA; Prof. Titular da UFC, Tutor do PET-Pedagogia.

1. Introdução

Atualmente a função de **avaliar** ocupa posição destacada nas atividades humanas dotadas de planejamento sistemático mínimo. Compreende-se, facilmente, o porquê. Num plano modesto ou de grande dimensão, a curto ou a longo prazo, sempre são visadas metas ou objetivos. Para toda ação planejada alcançar com êxito as suas metas é importante, por um lado, avaliar se está decorrendo conforme previsto e, por outro, avaliar se os resultados obtidos são de fato os pretendidos (Lemos, 1993).

Assim, é necessário verificar, ao longo do percurso, se estão ocorrendo erros ou desvios que irão impedir a obtenção do produto desejado. A avaliação ao longo do trajeto permite a intro-

1 - Estudo realizado no Laboratório de Pesquisas em Avaliação e Medida Psico-Educacional (LABPAM/UFC).
2 - Filiação Institucional: Professor do Curso de Mestrado em Avaliação Educacional (FACED/UFC-Brasil)
3 - Filiação Institucional: Professor do Curso de Mestrado em Avaliação Educacional (FACED/UFC-Brasil)

dução de correções ou novas estratégias que possam reconduzir o projeto à rota traçada *(Ribeiro, 1993)*.

No campo específico da Educação, a avaliação pode adotar uma diversidade de formas e dimensões de acordo com os seus objetivos. O quadro 1 apresenta esquematicamente as formas e dimensões que pode assumir a avaliação, de acordo com *Hadji (1994)*.

QUADRO 1: Formas e dimensões da avaliação

FORMAS	EXEMPLOS DE DIMENSÕES
Avaliação de instituições	1) Secretaria de Estado 2) Escola
Avaliação de projetos e programas	1) Projeto de reforma educacional 2) Currículo de um nível de ensino
Avaliação de serviços prestados por uma instituição	1) Aspectos administrativos 2) Atividades de ensino
Avaliação de grupos de intervenientes	1) Administradores 2) Professores e alunos
Avaliação de processos e produtos	1) Eficácia de métodos de ensino ou treinamento 2) Aprendizagem

A divisão apresentada no quadro 1 é meramente didática. Na realidade, as formas de avaliar não são mutuamente excludentes. Neste estudo será tratada a avaliação da aprendizagem, que é uma das dimensões da avaliação de processos e produtos.

2. Avaliação do Processo Ensino-Aprendizagem

Dentro do campo educacional, o conceito de **avaliação da aprendizagem** tem gerado inúmeros debates e continua a ser fruto de muitas polêmicas.

Em outra ocasião, foi defendida a idéia de que, malgrado as dificuldades que tal empreitada encerra, pode-se postular uma definição de avaliação que possibilite um grau maior de concordância entre os estudiosos do assunto *(Barreto, 1993)*. Parece que a conceituação apresentada por *James Popham (1977)* atende a este critério. Diz ele:

> *embora venham ocorrendo algumas diferenças de opinião através dos anos, a maioria dos educadores concebe a avaliação educacional como operação na qual a qualidade de uma iniciativa educacional é julgada. Em outras palavras, para a maior parte dos educadores o termo avaliação significa o julgamento do valor de uma iniciativa educacional.*

Tal modo de ver a avaliação exige, para que possa ser efetivada, que, para cada aspecto da realidade educacional a ser avaliado, sejam definidos padrões de comparação. E é justamente no julgamento comparativo entre a ação educativa e os padrões definidos que consiste a avaliação. Pode-se, portanto, afirmar que

> *uma maneira conveniente de pensar sobre avaliação é considerá-la como ato de aferir, através de comparação, (1) o resultado observado (dados de desempenho) de alguma iniciativa educacional com (2) um padrão ou critério de aceitabilidade pretendido (dados desejados) (Popham, 1977).*

Tyler (1981) por sua vez, ao se limitar à sala de aula, afirma que,

> *O processo de avaliação da aprendizagem consiste, essencialmente, em determinar em que medida os objetivos educacionais estão sendo realmente alcançados pelo programa planejado (p. 98).*

Analisando as inúmeras definições de avaliação, *Silva (1992)* revela a existência de alguns pontos comuns entre estas:

a) o termo *avaliação* difere semanticamente de *medida*, porém a inclui como condição indispensável à sua objetividade e precisão;

b) a avaliação realiza-se em função de objetivos claramente definidos;

c) a avaliação é um processo amplo, influenciado por diferentes aspectos da situação educacional;

d) a avaliação fornece informações úteis para a tomada de decisões com relação a alunos;

e) a avaliação é uma atividade que deve ser contínua, sistemática e científica.

A estas características deve-se acrescentar uma que, sem dúvida, se constitui no coroamento das ações avaliativas, muito bem expressa na definição de Popham mencionada na página ao lado: **a avaliação é um julgamento de valor.**

É claro que, para a avaliação do desempenho escolar em sala de aula, o professor pode e deve usar todas as informações ao seu dispor. Entretanto, nas avaliações envolvendo um número muito grande de alunos, existem limitações de ordem prática (restrições financeiras, dificuldades operacionais) que obrigam a utilização de testes objetivos que possibilitem, pelo menos, uma medida válida e precisa do rendimento.

Esta medida é fundamental para um processo educacional eficiente, pois indica em que grau os objetivos educacionais cognitivos foram alcançados e permite a classificação dos alunos. Por outro lado, possibilita aos professores a revisão dos métodos de ensino empregados e aos responsáveis pela política educacional e administradores de escolas a correção das distorções no planejamento e execução das atividades educacionais.

Para obter essa medida do rendimento escolar, a Secretaria de Educação e Cultura do Estado do Ceará (SEDUC), juntamente com o Departamento de Fundamentos da Educação da Faculdade de Educação da Universidade Federal do Ceará, aplicou na população das escolas

públicas de 27 municípios[4] testes referentes a critério para avaliação da aprendizagem nas disciplinas *português* e *matemática* (4ª e 8ª séries). O objetivo do presente estudo é descrever as fases da construção do teste de português para a 4ª série e o modelo teórico de respostas ao item (TRI) adotado para a determinação dos parâmetros métricos dos seus itens.

3. Considerações Gerais sobre a Teoria de Respostas ao Item (TRI)

A Teoria de Respostas ao Item (TRI) utiliza duas funções matemáticas para caracterizar os parâmetros métricos dos itens de um teste: a *logística* e a *normal padronizada*. Ambas fornecem informações sobre os parâmetros dos itens através das curvas características dos itens (CCI's). As CCI's não são iguais, apesar de compartilharem alguns aspectos na sua forma geral. Também é possível a obtenção de informações sobre o teste através da sua curva característica (CCT), que fornece o erro-padrão de medida, ou seja, a quantidade de erro referente à magnitude da variável medida, presente no teste.

As informações contidas nas CCI's a respeito dos parâmetros métricos dos itens, dependem do modelo teórico escolhido. O mais simples foi sugerido por G. Rasch em 1960; recebeu o nome de *modelo logístico de um parâmetro* e contém o presuposto de que a probabilidade de acerto de um item é influenciada pela sua dificuldade. Sua formulação matemática é: $P_1(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}}$, onde:

$P_1(\theta)$: probabilidade de acertar o item $_i$ para um determinado valor de θ;

θ: valor da variável medida;

b_i: índice de dificuldade do item $_i$;

e: base dos logaritmos neperianos (2,72);

D: constante de valor 1,7 (com este valor, a função logística aproxima-se notavelmente da curva normal padronizada).

O segundo tipo, denominado *modelo logístico de dois parâmetros*, foi apresentado por volta de 1968 por A. Birnbaum. Neste modelo, a probabilidade de acerto de um item é influenciada pela sua dificuldade e discriminação. Sua definição matemática é: $P_2(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$, onde $P_2(\theta)$, θ, b_i, e, D assumem o mesmo significado do modelo de um parâmetro. Sua diferença está no aparecimento, na sua formulação, do índice de discriminação do item a_i.

Por último, o *modelo logístico de três parâmetros*, também desenvolvido a partir dos trabalhos de A. Birnbaum, assume o entendimento de que a probabilidade

4 - Foram escolhidos pela SEDUC os municípios-sedes dos Conselhos Regionais de Desenvolvimento Educacional (CREDE's) e aqueles onde o ensino de 1º grau foi totalmente municipalizado.

de acerto de um item é influenciada pela sua dificuldade, discriminação e probabilidade de acerto ao acaso. Em termos matemáticos, o modelo é expresso por:

$$P_3(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

onde $P_3(\theta)$, θ, a_i, b_i e, D possuem o mesmo significado dos modelos aqui mencionados e c_i indica a probabilidade de acerto ao acaso.

Para a utilização dos modelos componentes da TRI há algumas etapas a serem seguidas e que devem ser caracterizadas associando-as ao processo avaliativo.

4 - Definição da Variável a ser Avaliada

A variável avaliada foi a aprendizagem dos alunos de escolas públicas do Estado do Ceará, da 4ª série do 1° grau, na disciplina *português*. Na definição da variável a ser mensurada, dois aspectos foram considerados: 1) sua conceituação, ou seja, o que se entendeu, no presente caso, como aprendizagem, e 2) sua definição operacional, isto é, quais as tarefas que permitem a avaliação do nível de aprendizagem dos alunos em português.

Nos limites da avaliação realizada, a aprendizagem em português foi definida como o domínio, pelos alunos, dos conteúdos cognitivos da proposta curricular elaborada pela SEDUC. Esta definição diminui a extensão do conceito de aprendizagem e aumenta sua compreensão facilitando, assim, sua definição operacional e posterior elaboração dos itens do teste.

A definição operacional foi realizada através da adoção da taxonomia dos objetivos educacionais proposta por Benjamim Bloom. Esta classificação supõe a existência de níveis diferenciados de aprendizagem, partindo de processos básicos (memorização, interpretação e compreensão) e indo até procedimentos mais complexos (análise, síntese e avaliação). Esses processos foram definidos e caracterizados *a priori*. Considerando a faixa etária dos respondentes e o seu desenvolvimento cognitivo, o teste englobou questões que procuravam avaliar os três níveis mais básicos de aprendizagem segundo a taxonomia adotada.

5. Elaboração dos Itens Destinados a Medir a Variável

Conforme afirmação de Costa (1995), *a tarefa de construir em larga escala bons itens depende do cumprimento de importantes fases*. A primeira dessas fases consistiu em decidir que disciplinas seriam objeto da avaliação e em que grau escolar. A segunda etapa constou da escolha de um especialista para o trabalho de elaboração de itens. Em seguida, organizou-se uma matriz ou tabela de especificação cuja função foi a de orientar o elaborador de itens quanto: a) aos processos cognitivos a serem avaliados, b) à consonância com os conteúdos curriculares propostos pela Secretaria de Educação e Cultura do Estado do Ceará, c) à distribuição dos itens segundo as faixas de dificuldade e d) ao número de itens. Nesta fase inicial de trabalho foram elaborados 35 itens de múltipla escolha com cinco alternativas de resposta.

6. Pré-Teste dos Itens

O principal objetivo dessa fase de trabalho foi a filtragem ou limpeza dos itens, sendo o procedimento mais pertinente a determinação do índice de discriminação através da correlação ponto-bisserial *(Muñiz Fernández, 1990)*. Para tanto, utilizou-se uma amostra de 100 alunos da 4ª série de uma escola não participante da avaliação, tendo como idade média 12,35 anos (DP=1,88) e moda relativa a 13 anos (23% dos alunos), com uma ligeira predominância do sexo feminino (57%).

Os resultados indicaram que oito itens obtiveram coeficientes de correlação inferiores a 0,20, ou seja, não discriminaram os sujeitos com níveis diferenciados de aprendizagem em português.

7. Organização do Instrumento Final e Aplicação

A forma final do teste de português ficou composta pelos 25 itens de mais elevados índices de discriminação. A etapa posterior consistiu na aplicação do referido instrumento na população estudada. Foram 17576 alunos da 4ª série do 1º grau, cuja idade média foi de 12, 27 anos (DP=1,94), sendo a maioria de 11 anos (23,2%) e do sexo feminino (53,8%).

8. Comprovação dos Pressupostos Teóricos da TRI

Os pressupostos referem-se a duas características que devem possuir os itens. A primeira está associada à unidimensionalidade, ou seja, o grupo de itens deve medir uma mesma variável. O método mais pertinente para a sua comprovação é a análise fatorial. Aplicada no grupo de 25 itens, a análise fatorial (método dos componentes principais) extraiu 4 fatores com valores próprios *(eigenvalues)* superiores a 1,0 e explicando juntos 31,5% da variância total dos escores, sendo que o 1º fator extraído explicou 16% (valor próprio=3,988), ou seja, explicou sozinho 50,8% da variância comum aos 4 fatores. Assim, pode-se afirmar que o teste possui elevada unidimensionalidade.

O segundo pressuposto é referente à independência local dos itens, isto é, que para um sujeito com determinado valor na variável unidimensional (θ) sua resposta a um item não vem influenciada por suas respostas a outros itens. Segundo *Hambleton, Swaminathan & Rogers (c1991)*, se a unidimensionalidade é cumprida, disto deriva, matematicamente, a independência local entre os itens, dado que os dois conceitos são equivalentes.

9. Estimação dos Parâmetros dos Itens

Para os dados disponíveis, o modelo que permitiu uma análise mais adequada foi o modelo logístico de dois parâmetros.

Objetivando a estimação dos parâmetros métricos dos 25 itens componentes do teste de português e a representação gráfica das suas curvas características (CCI's), utilizou-se o *software* BILOG

FOR WINDOWS (3.09). Os itens 10, 13, 16, 17, 18, 24 e 25 foram omitidos das análises posteriores por possuírem variância muito reduzida nas respostas que receberam. Os resultados são apresentados e comentados em anexo.

Do anexo 1 constam os gráficos representativos da *função de informação do teste* e do *erro padrão de medida*. O anexo 2 apresenta, a título de ilustração, dois gráficos com diferentes informações sobre as CCI's de 5 itens. No primeiro tipo ($P_2(\theta)$, isto é, a função de resposta ao item e a percentagem de respostas corretas esperadas) há os parâmetros métricos, isto é, a discriminação **a** e a dificuldade **b**. Também consta a distribuição teórica (gráfico de $P_2(\theta)$, isto é, curva de linha contínua) e as freqüências observadas (pontos em torno ou sobre a curva de $P_2(\theta)$) para determinadas magnitudes de θ.

No segundo tipo de gráfico ($P_2(\theta)$ e função de informação do item) há para cada item a sua função de informação (em linhas tracejadas), ou seja, a magnitude do θ em que, ao se utilizar o item, se obtém a maior quantidade de informação e se comete a menor quantidade de erros de medida.

10. Comprovação do Ajuste do Modelo aos Dados

Há vários procedimentos para estimar o ajuste do modelo aos dados, entre os quais destacam-se o teste do qui-quadrado, a análise dos resíduos e a comparação das distribuições das pontuações. No presente caso, utilizou-se o teste do qui-quadrado. Assim, as freqüências observadas foram comparadas com as freqüências teóricas, ou seja, a distribuição empírica foi comparada com a distribuição teorizada pelo modelo adotado, ressaltando-se o fato de que tal comparação é feita item por item.

Os resultados revelaram que as distribuições empíricas dos itens 1, 2, 3, 4, 7, 8, 9, 11, 12 e 14 se ajustaram às distribuições teóricas, visto que os valores dos qui-quadrados não foram significativos adotando-se $p<0,05$, isto é, os valores observados a partir dos resultados do teste na sua função de informação não diferem dos valores de $P_2(\theta)$. Entretanto os itens 11, 12 e 14 por apresentarem valores de **b** muito altos (respectivamente 3,9; 2,8 e 5,0) não têm praticamente nenhum poder de discriminação e deverão ser eliminados de testes futuros.

11. Considerações Finais

As discussões sobre avaliação têm se intensificado sobremaneira, no meio educacional brasileiro, nos últimos dez anos. Esta preocupação ganhou maior amplitude, a partir de 1988, com a criação do Sistema Nacional de Avaliação das Escolas Públicas do 1° grau (SAEP), denominado, a partir de 1991, de Sistema Nacional de Avaliação da Educação Básica (SAEB).

Esta fase de experiências e discussão localizadas teve como resultado abalar um pouco as resistências à implantação de programas de avaliação em cada *escola*

e nos sistemas escolares. Entretanto, é importante ressaltar a necessidade de se evitar dois obstáculos que comprometerão os esforços de pessoal e instituições.

O primeiro é o modismo. A experiência registra vários surtos desta *doença*. Parece que os educadores são particularmente suscetíveis de *contaminação* pelo vírus da novidade. É necessário, portanto, uma *vacina* contra este mal e que a avaliação seja encarada como uma necessidade imediata para a melhoria da educação nacional.

O segundo embaraço é a superficialidade, o faz-de-conta, a capacidade que temos para disfarçar a realidade. Os educadores são mestres em mascarar o real e dar aparência de mudanças a ações que visam a manter o *status quo*. A eles se aplica com muita propriedade o adágio francês: ***plus ça change, plus ça c'est la même chose***.

Para que estas barreiras sejam superadas algumas precauções devem ser tomadas. Uma delas refere-se à necessidade de se estabelecer com a maior precisão e clareza possível o conceito de avaliação e a sua distinção quanto à técnica de medida *(Barreto, 1993)*.

Deve ser dito que as medidas efetuadas num processo de avaliação podem apenas oferecer subsídios para o ato de julgar, prerrogativa do avaliador. Entretanto, quando utilizadas, as medidas hão de ser feitas com critérios e os instrumentos empregados têm de possuir características indispensáveis.

Andriola (1995 e 1996) destaca o fato de que os instrumentos de medida, tanto em psicologia quanto em educação, necessitam de parâmetros métricos determinados através de modelos teóricos. Segundo o Autor, sem a determinação empírica dos parâmetros métricos inerentes aos itens e ao instrumento de medida, a confiabilidade dos resultados da avaliação ficará comprometida. No presente caso, o modelo teórico adotado para a determinação empírica dos parâmetros métricos dos itens componentes de um instrumento de medida foi a Teoria de Respostas ao Item (TRI).

Destaque-se a verdade de que as grandes contribuições da TRI na área da avaliação psicológica e educacional são as de proporcionar: a) a organização de bancos de itens estritamente calibrados, isto é, itens cujos parâmetros métricos estejam determinados, e b) a obtenção de medições invariantes em relação aos instrumentos utilizados e aos sujeitos implicados *(Fletcher, 1994)*. De posse de um grupo (banco) de itens cujos parâmetros métricos e suas curvas características tenham sido determinadas, os responsáveis pela avaliação poderão combinar os itens de modo a formar diferentes testes congruentes com os seus objetivos. Por exemplo, se é objetivo diferenciar os respondentes com elevada aprendizagem ($\theta > 1{,}5$), utilizar-se-ão os itens que discriminem magnitudes de θ superiores a 1,5! Esse procedimento exemplificado demonstra outra vantagem que a TRI oferece: a possibilidade de o avaliador adaptar os testes ao nível de conhecimento ou

aptidão dos respondentes, permitindo, assim, avaliações mais exaustivas e rigorosas em função das características dos sujeitos, ou seja, em função da magnitude do θ dos respondentes.

Com os itens aqui apresentados, cujos parâmetros métricos foram determinados através do modelo logístico de dois parâmetros, há possibilidade de que sejam utilizados por outros avaliadores, em outras regiões do país e com outros respondentes. Afinal, como afirma a TRI, há uma relação funcional entre a variável que os itens medem (θ) e a probabilidade de acertá-los.

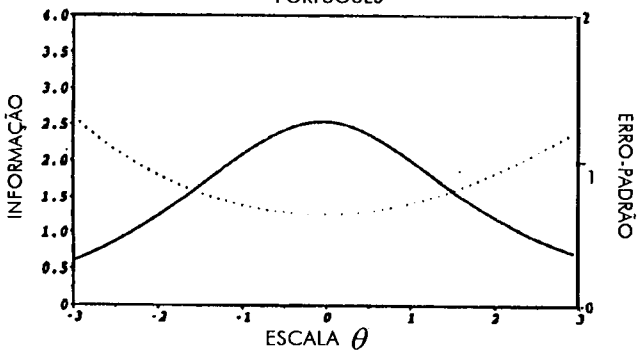
Para finalizar, há de ser ressaltado que a TRI possui alguns problemas que são motivos de controvérsias e polêmicas. Um desses refere-se ao pressuposto da unidimensionalidade dos itens e às limitações impostas pelos métodos utilizados para sua comprovação. É necessária, portanto, a proposição de novos procedimentos para o teste desse pressuposto essencial à TRI.

ANEXO 1

Curva característica do teste (CCT) de Português (4ª série)

CURVA CARACTERÍSTICA DO TESTE (FUNÇÃO DE INFORMAÇÃO E DE ERRO-PADRÃO DE MEDIDA)

PORTUGUÊS



Este gráfico indica o grau de informação do teste (curva com linha contínua) e do seu erro-padrão de medida (curva com linha tracejada). Observe que a variável latente θ, isto é, a aprendizagem em português, tem seus valores padronizados marcados no eixo das abcissas, no intervalo (-3, +3). No eixo das ordenadas aparecem o grau de informação do teste e do seu erro-padrão de medida.

A análise do gráfico permite afirmar:

1) o teste é de dificuldade média;

2) para valores de θ situados entre -1,5 e +1,5 o teste proporciona informação máxima. Isto significa que para os alunos que na distribuição da variável latente, aprendizagem em português, estão situados entre 1,5 desvios-padrões abaixo da média e 1,5 desvios-padrões acima da

média, o teste é preciso na sua medida;

3) neste mesmo intervalo, o erro-padrão de medida S_e assume seus menores valores. Pela leitura visual do gráfico S_e varia entre 0,67 e 0,8, aproximadamente;

4) para magnitudes de θ diferentes dos valores mencionados (-1,5 a +1,5) o teste perde rapidamente sua precisão e aumenta, conseqüentemente, o seu erro-padrão de medida.

ANEXO 2

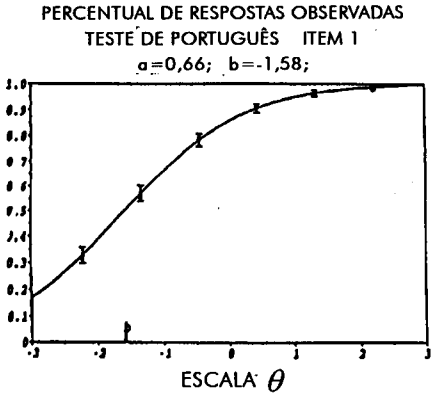
Curvas Características dos Itens (CCI's)

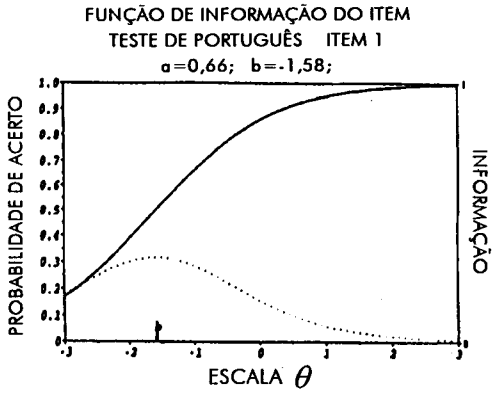
São apresentadas duas curvas características dos itens informando:

a) o percentual de respostas observadas (pontos em torno ou sobre a curva teórica),

b) a sua função de informação (curva com linha tracejada).

Ambas se contrapõem ao gráfico de $P_2(\theta)$ (modelo logístico de dois parâmetros). Será feito um comentário para as CCI's do item 1.

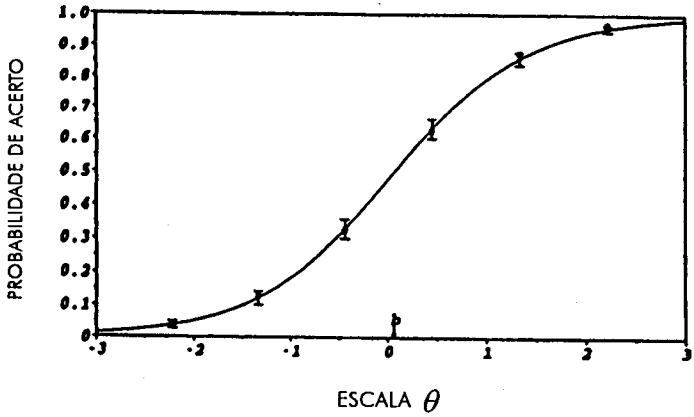




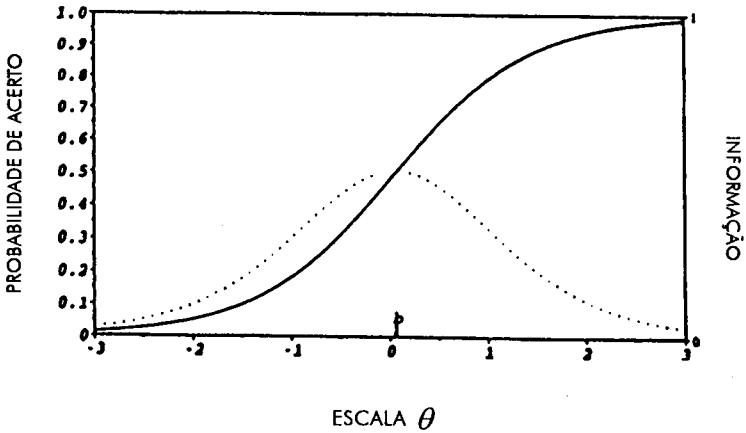
A análise das características métricas nesse estudo revelou, que os itens com a>0,5, são discriminativos para os valores assumidos por θ no intervalo de +1,7 e -1,7 (Escala do θ) a partir do valor de b. Assim, por exemplo, o item 1 assume para o parâmetro b o valor de -1,58, estando a magnitude de θ na qual mais precisamente discrimina compreendida entre os valores +0,12 e -3,28 na escala de θ. É exatamente neste intervalo que a função de informação do item é máxima (ver gráfico correspondente do item 1).

Em seguida, para a análise do leitor, são mostradas as CCI's dos itens 3, 4, 11 e 12.

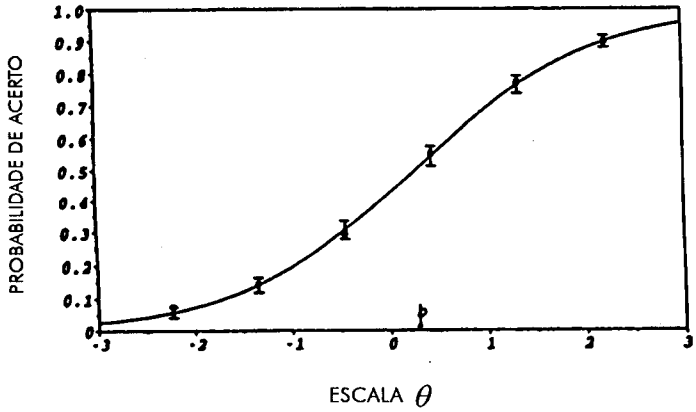
PERCENTUAL DE RESPOSTAS OBSERVADAS
TESTE DE PORTUGUÊS ITEM 3
a=0,84; b=-0,06;



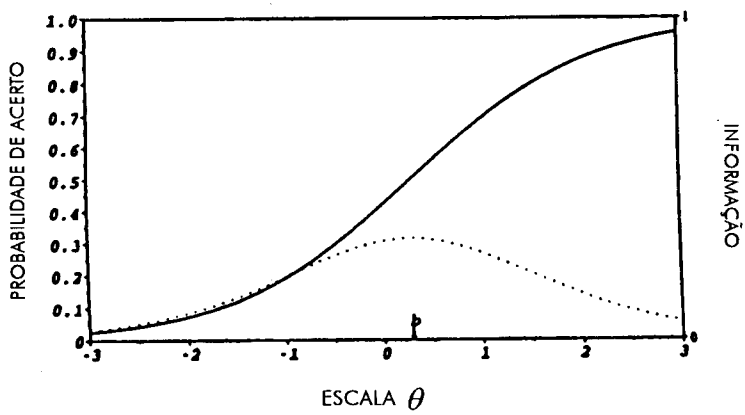
FUNÇÃO DE INFORMAÇÃO DO ITEM
TESTE DE PORTUGUÊS ITEM 3
a=0,84; b=-0,06;

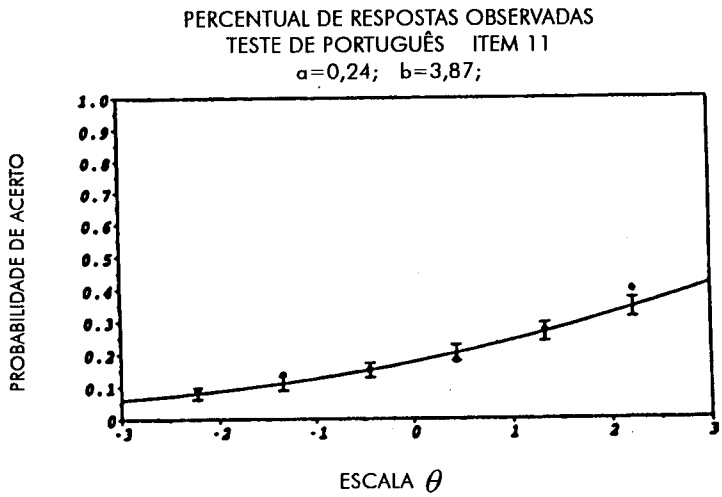


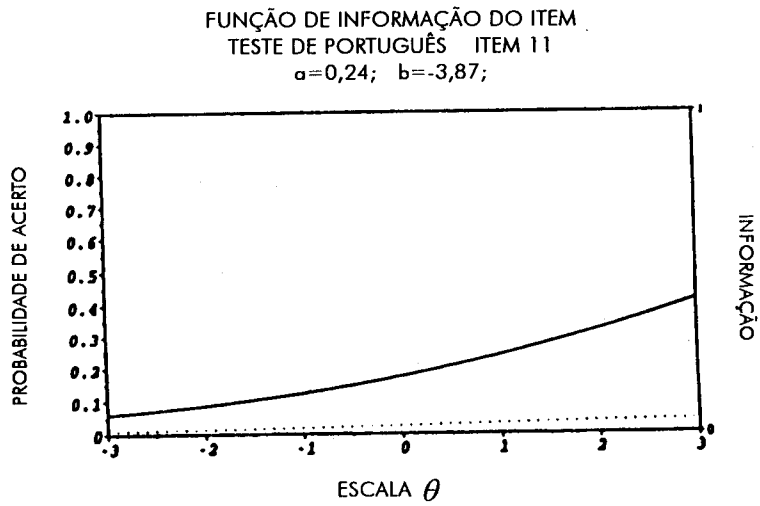
PERCENTUAL DE RESPOSTAS OBSERVADAS
TESTE DE PORTUGUÊS ITEM 4
a=0,66; b=-0,29;



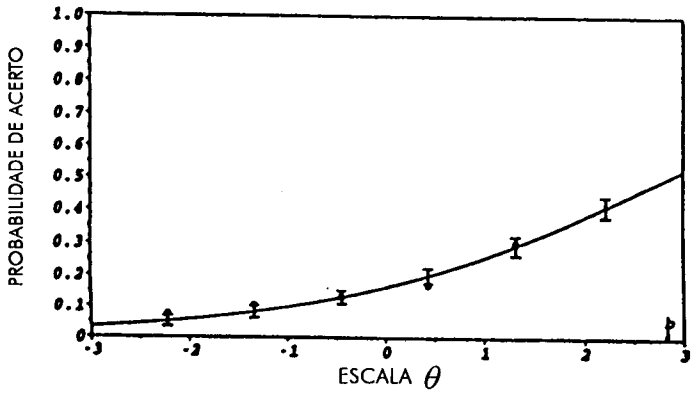
FUNÇÃO DE INFORMAÇÃO DO ITEM
TESTE DE PORTUGUÊS ITEM 4
a=0,66; b=-0,29;



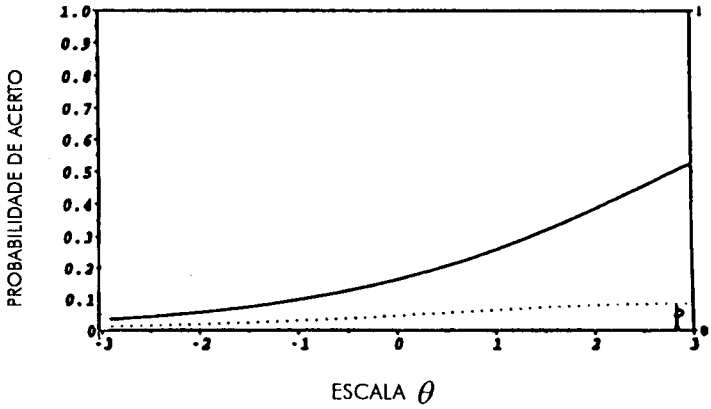

PERCENTUAL DE RESPOSTAS OBSERVADAS
TESTE DE PORTUGUÊS ITEM 11
a=0,24; b=3,87;
PROBABILIDADE DE ACERTO
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-3
-2
-1
0
1
2
3
ESCALA θ


FUNÇÃO DE INFORMAÇÃO DO ITEM
TESTE DE PORTUGUÊS ITEM 11
a=0,24; b=-3,87;
PROBABILIDADE DE ACERTO
INFORMAÇÃO
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-3
-2
-1
0
1
2
3
ESCALA θ

PERCENTUAL DE RESPOSTAS OBSERVADAS
TESTE DE PORTUGUÊS ITEM 12
a=0,34; b=2,84;



FUNÇÃO DE INFORMAÇÃO DO ITEM
TESTE DE PORTUGUÊS ITEM 12
a=0,34; b=-2,84;



ABSTRACT

This research is the result of evaluation of learning process of students in public schools in the state of Ceará. It was carried out towards the end of 1996. The objective was to estimate the metric parameters of the items components of a portuguese test destinated for the students of the 4th year of primary school, using the Item Response Theory (IRT). Analysing the answers of 17576 students with an average age of 12,27(DP=1,94), the dificulty and discrimination of the 25 items on the portuguese test were estimated by using the logistic model of two parameters. The items caracteristic curves (ICC's) and those of the test (TCC) were also estimated.

RESUMEN

Este estudio es resultado del proceso de evaluación del aprendizaje de los alumnos de escuelas públicas del Estado do Ceará, realizado a finales de 1996. El objetivo fue la estimación de los parametros métricos de los items que componen el test de portugués destinada a los alumnos del 4° curso, a través de la Teoría de Respuestas al Item(TRI). Utilizándose las respuestas de 17576 alumnos, cuya edad media fue 12,27 años (DP=1,94), se estimó la dificultad y la discriminación de los 25 items que componen el test de portugués a través del modelo logístico de dos parametros, determinandose, también, las curvas características de los items (CCI's) y del test (CCT).

Referências Bibliográficas

ANDRIOLA, W. B. Avaliação psicológica no Brasil: considerações a respeito da formação dos psicólogos e dos instrumentos utilizados. *Psique*. Belo Horizonte, v.8, n.1, p.98-108, maio 1996.

_____. Os testes psicológicos no Brasil: problemas, pesquisas e perspectivas para o futuro. In: ALMEIDA, L. S. , RIBEIRO, I. S. (Org.) *Avaliação psicológica*: formas e contextos. 4.ed. Lisboa: Texto Editora, 1995. v.3 p.77-82.

BARRETO, J. A. E. Avaliação: mitos e armadilhas. *Ensaio: Avaliação e Políticas Públicas em Educação*. v.1, n.1, p.43-54, out./dez. 1993.

COSTA, S. F. Matriz de especificações: suporte operacional de uma avaliação. *Estudos em Avaliação Educacional*, São Paulo, n.12, p.59-65, jul./dez, 1995.

FLETCHER, P. R. Teoria de resposta ao item: medidas invariantes do desempenho escolar. *Ensaio: Avaliação e Políticas Públicas em Educação*, v.1, n.2, p.21-7, jan./mar. 1994.

HADJI, C. *A avaliação, regras do jogo*: das intenções aos instrumentos. Tradução por: Julia Lopes Ferreira, José Manuel Claudio. 4.ed. Porto (Port.): Porto, 1994. 190p. (Ciências da Educação; 15) Tradução de: L'evolution, regle du jeu)

HAMBLETON, R. K., SWAMINATHAN, H., ROGERS, H. J. *Fundamentals of item response theory*. Newbury Park: Sage, c1991. x, 174p. (Measurement methods for the social sciences series; v.3)

LEMOS, V. *O critério do sucesso*: técnicas de avaliação de aprendizagem. Lisboa: Texto Editora, 1993.

MUÑIZ FERNÁNDEZ, J. *Teoría de respuesta a los ítems*: un nuevo enfoque en la evolución psicológica y educativa. Madrid: Pirámide, 1990. (Colección Psicología)

POPHAM, W. J. *Manual de avaliação*: regras para o avaliador educacional. Tradução por: Cecília L. da Rocha Bastos. Petrópolis (RJ): Vozes, 1977. 90p. Tradução de: Educational evaluation.

RIBEIRO, L. C. *Avaliação da aprendizagem*. Lisboa: Texto Editora, 1993.

SILVA, C. S. *Medidas e avaliação em Educação*. Petrópolis (RJ): Vozes, 1992.

TYLER, L. E. *Testes e medidas*. Tradução por: Alvaro Cabral. 3.ed. Rio de Janeiro: Zahar, 1981. 135p. (Curso de pedagogia moderna) Tradução de: Tests and measurements.