

A Teoria da Resposta ao Item: Medidas Invariantes do Desempenho Escolar *

Philip R. Fletcher

A sociedade democrática, em sua essência, permite um alto grau de descentralização e de autonomia, que, por sua vez, permite uma melhor adaptação às condições locais e um melhor entrosamento com a comunidade.

Afinal, este é o propósito da democracia. Contudo, a sociedade democrática não é por isso permissiva ou omissa. Rege controles compatíveis com uma administração descentralizada e autônoma. Dentre estes controles, quero ressaltar a importância do que chamamos de "accountability" em inglês. De certo, não tem nada a ver com a área econômica ou contábil. É um termo muito mais relacionado com a organização administrativa e política. Neste particular, é talvez significativo que não exista tradução para o termo em português.

Pode-se traduzir "accountability" pela frase "prestar contas", embora perca-se uma parte de seu significado substantivo ao fazer a tradução. Vejamos, então, que "accountability" implica cinco coisas:

1. Encargo ou responsabilidade;
2. Informações sobre o desempenho realizado;

3. Critérios de desempenho esperado;

4. Juízo que relaciona informações e critérios num caso particular; e o que distingue a "accountability" da simples prestação de contas.

Philip R. Fletcher

*Ph.D em Educação, pela
Universidade de
Stanford, Califórnia - USA;
Pesquisador Sênior da
Fundação Carlos Chagas.*

As informações, critérios, juízo e retribuição são de domínio público e, portanto, conhecíveis por cidadãos que participam do processo de avaliação no decorrer do cotidiano. Assim se diz que o Presidente terá que prestar contas na próxima eleição, que seu desempenho será avaliado e recompensado através do voto nas próximas eleições; da mesma forma que a

Diretores, Professores e Alunos de escolas é permitido um grau adicional de liberdade em suas atividades habituais, em troca da natural habitual e conseqüente prestação de contas.

Dessa maneira, aprende-se a avaliar e a ser avaliado na escola como ato de iniciação à cidadania. É na escola que aprendemos a ganhar a recompensa devida. Mais tarde, na vida em sociedade, ninguém garante o pão de nosso dia, a imunidade ou a estabilidade, a não ser através da justa recompensa, dispensada com a periódica prestação de contas em

(*) Palestra proferida no "Simpósio Nacional sobre Avaliação Educacional: uma reflexão crítica". Realizado pela Fundação CESGRANRIO, Rio de Janeiro, Set/Out 1993.

ato público.

Com a "accountability", a sociedade se liberta das pressões autoritárias da vigilância administrativa externa e da intromissão alheia indevida. Isto porque a retribuição prometida pela "accountability" permanece iminente e será efetuada apenas eventualmente. Assim, as recompensas de um desempenho acima da média virão apenas a longo prazo, possivelmente para coroar muitos anos de bom desempenho profissional ou acadêmico. Da mesma maneira, a devida recompensa para um desempenho deficiente não será imediata, mas, eventualmente, poderia realizar-se após muitos anos de insucesso comprovado e quando as possibilidades de melhora já pareçam esgotadas.

A iminência da retribuição é suficiente como controle administrativo, na grande maioria das vezes porque normalmente as pessoas conseguem tirar as devidas conclusões quando informadas sobre seu desempenho, o que de fato elimina uma das maiores fontes de sua insegurança. Dessa forma, a "accountability" não é nada inquisitorial, o que vai de encontro a uma preocupação do professorado no ensino. Pelo contrário, ela permite ampla margem de autonomia e liberdade de ação, abrindo oportunidades para a expressão de uma diversidade de habilidades e talentos individuais mais coerentes com as condições locais.

Na essência da "accountability" temos encargos, informações, critérios, juízo e recompensas. Cabe à avaliação um papel de destaque nesse processo. Sobretudo, no fornecimento de informações que servem para julgar o desempenho realizado. São estas as informações que regem as negociações, as lutas e as acomodações em milhares de decisões individuais e coletivas e que, enfim, consubstanciam os processos de mudança em sociedades democráticas.

Faltam, no entanto, placas de sinalização, nos horizontes brasileiros, para orientar a mudança educacional. Não existem termômetros da educação brasileira. É como nave-

gar pelas mazelas da economia brasileira sem as habituais referências do IGP-M ou do câmbio para orientar nossas decisões econômicas.

Amyr Klimk não se atreve a navegar em águas turvas sem noção de direção, nem da distância. Através de um sistema de satélites e um microcomputador de mão, ele sabe sua posição dentro de apenas cinco metros em qualquer mar do mundo. Embora não deveríamos é o que fazemos todos os dias na educação nacional, com técnicas de avaliação regidas pelo autodidatismo, o amadorismo e a habitual improvisação. Não dispomos de escalas de escolaridade para os diversos assuntos em que as crianças são avaliadas. A rigor, não há como comparar avaliações feitas em diferentes ocasiões por diferentes professores. Na ausência dessa avaliação, perpetuamos as reprovações em massa e o desprezo sistemático com que ignoramos o que, de fato, se aprende e não se aprende nas escolas brasileiras, fraudando o conceito da cidadania logo na porta da escola.

"Navegar é preciso". No entanto, estamos soltos num mar sem bússola, com termômetros de borracha e réguas métricas que dependem do comprimento dos objetos medidos. Faltam às técnicas de avaliação procedimentos de medição objetivos, que permitiriam a uma sociedade democrática navegar em águas turvas.

Pratica-se, no Brasil, uma técnica de testes e medidas ultrapassada, obsoleta hoje em pelo menos vinte anos. Os técnicos da área de avaliação praticam o que se tornou a chamar de "Teoria Clássica" ou de "Teoria dos Escores Verdadeiros". Na prática, o que se faz é resumir o desempenho num teste pelo somatório dos acertos dos itens:

$$X_i = \sum_{j=1}^n w_j x_{ij}$$

Como a teoria clássica não consegue determinar a importância a ser atribuída a cada item, pressupõe-se que w_j seja igual ao constante $1/n$ com o resultado de X_i , se tornar uma

simples proporção de acertos. O resultado obtido depende exclusivamente do conteúdo da prova. Muda a dificuldade da prova, muda seu resultado. Torna-se precária a tentativa de representar os desempenhos das pessoas sobre essas bases sempre flutuantes. Uma situação nada satisfatória para quem pretende atravessar o Atlântico.

Surge, então, uma nova teoria que incorpora os princípios da teoria clássica e ainda a estende. A nova teoria chama-se "Teoria da Resposta ao Item", "Teoria de Características Latentes", ou então, "Teoria das Curvas de Característica".

A nova teoria conquista um crescente número de adeptos em psicologia e educação mundo afora porque oferece recursos mais flexíveis e eficazes na confecção, análise e apresentação dos resultados de prova que quaisquer outros recursos equivalentes, derivados a partir da teoria clássica. O cerne de seu potencial se encontra na relação que estabelece entre as características dos itens e as características operacionais da prova composta desses itens. Em sua essência, esta relação é invariável e permanente, não dependendo do número de itens da prova, do subconjunto de itens selecionados ou das habilidades das pessoas que participam da prova. Quaisquer outros resultados comparáveis advindos da teoria clássica servem apenas para testes hipotéticos de uma infinidade de itens aplicados a uma única população padrão.

O item se torna a nova unidade de análise de preferência, ao contrário da prova em seu conjunto, que unidade de análise da teoria clássica. Ao escolher o item, ou pequenos conjuntos de itens de mesma índole, como unidade de análise e peça de reposição intercambiável na elaboração das provas e na apresentação de seus resultados, a teoria da resposta ao item incentivou inúmeras inovações na prática da aplicação das provas, especialmente no desenvolvimento de bancos de itens e testes adaptivos. Os benefícios dos bancos de itens incluem a redução apreciável no tempo e custo da produção de provas

operacionais de alta qualidade. Testes adaptivos, na forma de instrumentos de papel e lápis usados em provas de duas etapas ou na forma de provas adaptivas administradas por computador, permitem que se reduza à metade ou a um terço o número de itens que seriam necessários numa prova convencional da mesma precisão.

De grande importância para programas de avaliação e acompanhamento a longo prazo é o potencial que a Teoria da Resposta ao Item proporciona para retirar e repor itens de uma prova operacional, sem alterar a interpretação da escala usada na apresentação de seus resultados. Tendo em vista que os escores de escala obtidos com a teoria da resposta ao item são funções dos parâmetros dos itens, a técnica de elaboração dos resultados automaticamente absorve quaisquer diferenças nas características dos itens que compõem a prova, inclusive sua dificuldade ou poder discriminatório diferentes. Desta forma, não há mais necessidade de procurar itens de reposição equivalentes aos itens aposentados ou de um estudo de reconciliação para tornar equivalentes as diferentes formas de uma prova, assim como seria o caso na teoria clássica. A teoria da resposta ao item permite que se desenvolvam testes comparáveis a partir de um grande acervo de itens e, ainda, prever suas qualidades técnicas antes de serem administrados, sem recorrer a um pré-teste dos instrumentos.

Outra característica própria da teoria da resposta ao item é sua capacidade de colocar itens e pessoas na mesma escala de desempenho. As pessoas são mais hábeis ou menos hábeis e os itens são mais difíceis ou menos difíceis na mesma escala de uma característica latente subjacente. O modelo de resposta que fundamenta a teoria da resposta ao item permite ao analista prever a probabilidade de acerto de uma pessoa com determinada habilidade representada pelo resultado da prova. Isso permite que resultados sejam referenciados a conteúdos. Itens representativos de pessoas ; que com determinado nível de habilidade conseguem acertar em 50 ou 80% das

vezes, revelam o significado dos resultados em termos do conteúdo das tarefas por elas dominadas.

No entanto, talvez o aspecto mais importante da nova teoria é a promessa de fornecer medidas invariantes do desempenho cognitivo, que não dependem dos itens que compõem a prova ou das pessoas investigadas na amostra. As características que distinguem uma escala da teoria da resposta ao item de seu equivalente clássico incluem a invariância dos parâmetros de itens em populações diversas e o potencial de relacionar novos itens a uma escala métrica fixa pré-estabelecida. A teoria da resposta ao item permite que se compare resultados de provas de dificuldade variável e ainda relacionar os resultados na mesma escala de desempenho cognitivo. A escala é fixa e permanece sempre uniforme e constante. Isso viabiliza o monitoramento das tendências cognitivas em indivíduos, série a série, e em escolas como instituições, ano a ano.

A realização do potencial da teoria da resposta ao item depende da estimação dos parâmetros de itens e de pessoas. Dentro de limites, qualquer população serve para calibrar os itens. A calibração fornece a cada item, parâmetros que caracterizam suas qualidades técnicas, independentes da população investigada. Esses parâmetros representam a dificuldade, o poder discriminatório e a porcentagem de acertos casuais característicos do item. Sendo invariantes, eles não dependem da amostra selecionada para fins de calibração. Sendo invariantes, podem ser aplicados a qualquer outra população, proporcionando resultados na mesma escala de habilidade. Isso permite que se substitua qualquer conjunto de itens, ou seja, que se substitua esta ou qualquer outra prova, permanecendo inalterados os resultados.

Apresento a seguir gráficos que mostram a possibilidade de desenvolver medidas invariantes do desempenho escolar no Brasil. O Gráfico 1 mostra a parametrização de um único item de prova tirado de um recente estudo do ciclo básico em Minas Gerais.

A linha mais grossa, em forma de "S", representa o modelo de resposta idealizado pela teoria de resposta ao item. As linhas finas tangenciais à curva representam, respectivamente:

1. O parâmetro b de dificuldade, representado pela linha reta vertical, que desloca a curva para a direita ou esquerda, dependendo da dificuldade do item;

2. O parâmetro a de poder discriminatório, representado pela linha tangencial angular, que altera o declive da curva, girando em torno de seu ponto de inflexão, dependendo do poder discriminatório do item; e

3. O parâmetro c de acerto casual, representado pela linha reta horizontal, que define o limite inferior ou assíntota da curva, uma porcentagem mínima, que sobe ou desce de acordo com as possibilidades de acerto do item quando não se conhece a resposta correta com certeza.

Compara-se essa idealização matemática com os pontos experimentais em torno da linha curvilínea, o que revela uma consistência bastante satisfatória entre a previsão teórica e a realidade empírica. As duas tendências teórica e empírica representam o aumento da probabilidade de acerto do item quando a habilidade das pessoas sobe. Os pontos experimentais representam grupos de pessoas de habilidade homogênea, encontradas em pequenas faixas ao longo da escala de habilidade. Cada ponto representa a proporção de acertos alcançada pelo grupo. A posição correspondente na curva seria a proporção de acertos esperada.

No caso particular do Item 6 da Prova de Matemática, o Gráfico mostra que essa pergunta se situa 0,752 desvio padrão acima da média da população, o que sugere que a pergunta seja apenas um pouco difícil para a população escolar de Minas Gerais como um todo. Seu poder discriminatório em torno desse nível de dificuldade é bastante satisfatório e a probabilidade de um acerto

sem saber a resposta correta com certeza é mínima, permanecendo abaixo dos 5%. No conjunto, tudo indica que o item será bastante útil na avaliação do desempenho dos alunos do ciclo básico.

Com a parametrização de vários itens do mesmo gênero, torna-se possível estimar parâmetros de habilidade θ (theta) para as pessoas que participam da prova. Isso leva em conta a configuração particular das res-

Rendimento do Aluno do Ciclo Básico de Alfabetização Prova de Matemática - Item 6

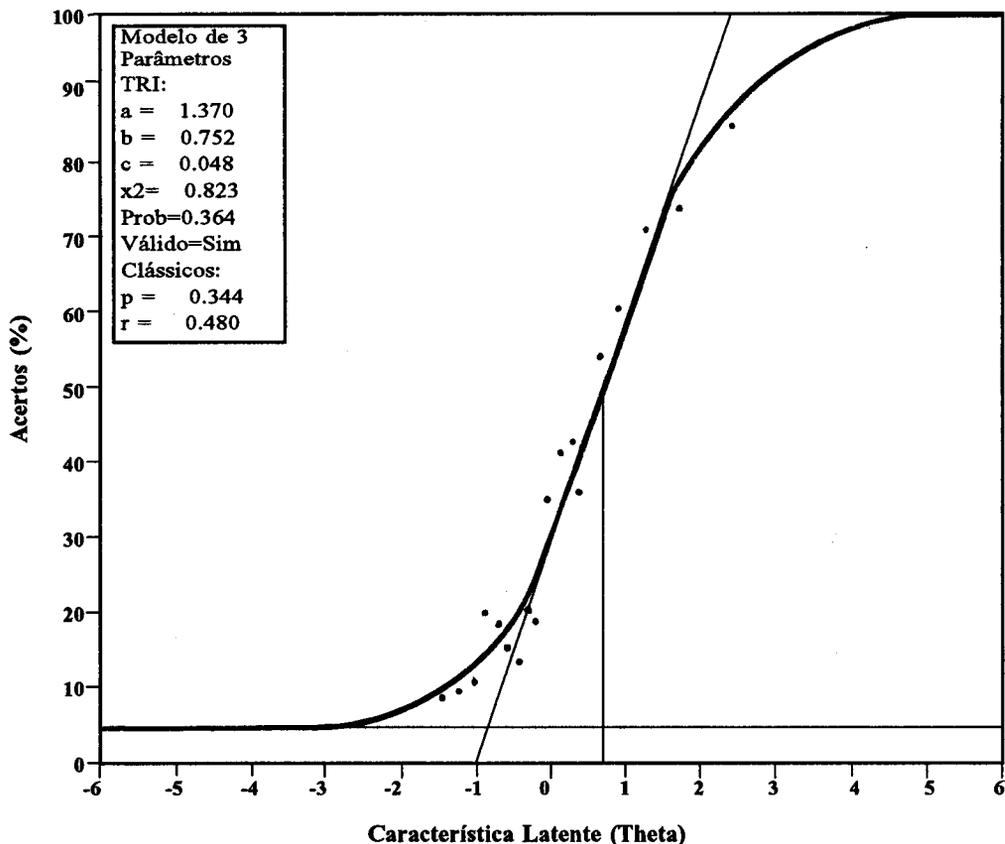


Gráfico 1

A Estimação dos Parâmetros do Modelo de Resposta de um Único Item de Prova

postas do indivíduo e a parametrização dos itens. Nesse caso, pessoas com percentuais de acerto iguais nem sempre recebem idênticas estimativas de habilidade. A teoria da resposta ao item sabe distinguir a importância relativa de cada item na composição do resultado final, representado por θ .

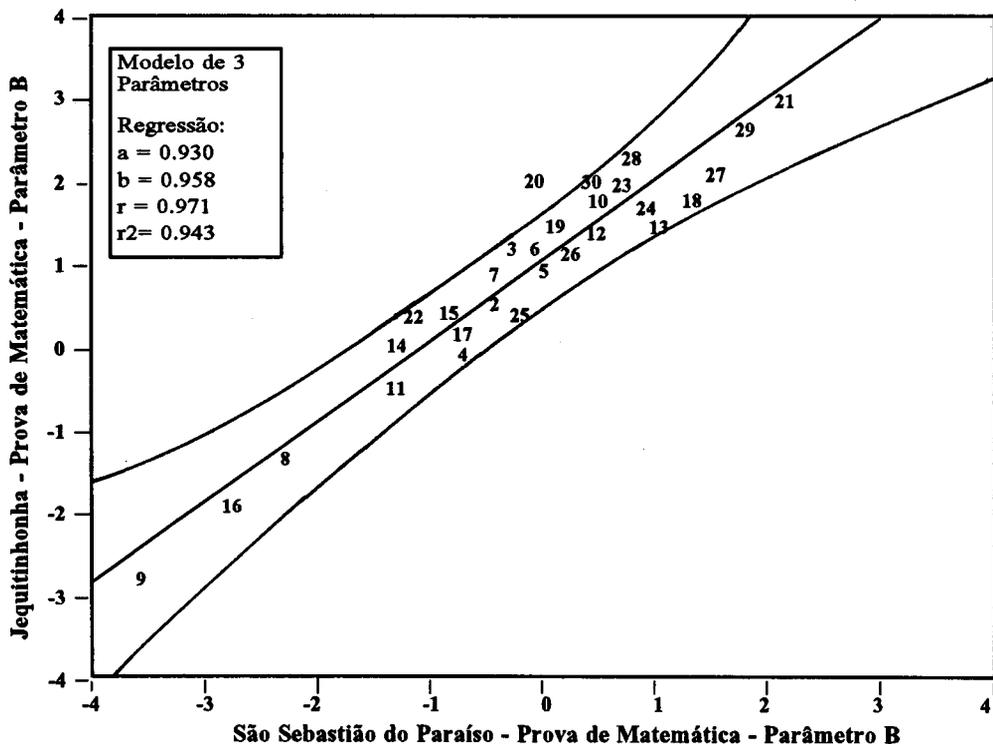
No entanto, tudo depende da parametrização dos itens. Pergunta-se, então, se esta parametrização é realmente invariante. Propõe-se examinar esta questão comparando dois extremos da distribuição de alunos em Minas Gerais. No Gráfico 2, compara-se o parâmetro b de dificuldade em duas calibrações independentes. No eixo horizontal, encontram-se os alunos do Diretório Regional de Ensino (DRE) de São Sebastião do Paraíso, que tiveram os melhores resultados em matemática de todo o Estado de Minas Gerais. No eixo vertical, encontram-se alunos do DRE do Vale

do Jequitinhonha, uma das regiões mais pobres do Estado.

A Teoria da Resposta ao Item promete que as duas calibrações independentes sejam iguais nas duas populações, com apenas uma transformação de origem e de escala. Se isto for verdade, então os parâmetros b , nos dois casos, devem cair ao longo de uma linha reta.

O intercepto a da linha de regressão, que relaciona a dificuldade dos itens nas duas populações de alunos, mostra que um aluno do Jequitinhonha precisa encontrar-se num desvio padrão acima da média dessa população para alcançar o mesmo nível de desempenho que um aluno médio em São Sebastião do Paraíso. Evidentemente, os alunos do Paraíso têm um desempenho superior em matemática, e por uma margem de magnitude apreciável.

Rendimento do Aluno do Ciclo Básico de Alfabetização



No entanto, o aspecto mais interessante do Gráfico é a nitidez da tendência linear das dificuldades dos itens, representada por um coeficiente de correlação Pearson de $r = 0,971$. A coeficiente de determinação de $r^2 = 0,943$ mostra que 94,3% da variação nas dificuldades de item é comum aos alunos do Jequitinhonha e de São Sebastião do Paraíso. Esses valores são inusitadamente altos para as ciências sociais. Eles representam uma concordância linear quase perfeita nas dificuldades *relativas* desses itens. Essa relação linear quase que perfeita é a base da invariância dos parâmetros da Teoria da Resposta ao Item, o princípio que permite a comparação dos resultados obtidos em provas de conteúdo equivalentes mas diferentes, ano a ano.

A teoria da resposta ao item promete modernas bússolas e régua métricas, que não as de borracha. As novas técnicas de avali-

ação em psicometria permitem aferir o estado cognitivo da população com uma nova objetividade e precisão, relacionadas a objetivos fixos bem definidos, ano a ano. São estas as informações que orientam as decisões educacionais em sociedades democráticas mundo afora. Ao saber se a escola vai bem ou vai mal, a sociedade, assim informada, consegue definir sua posição a respeito e cobrar ações que contribuem para o progressivo aperfeiçoamento do sistema. São estas as informações que permitem sua navegação em águas turvas.

Sem a avaliação objetiva do ensino, a nova sociedade que se democratiza corre o risco, vez por outra, de tornar-se saudosista da tutela e falsas pretensões dos antigos planos quinquenais de governo. Mas os percalços desse caminho já não enganam mais a ninguém.

RESUMO

A democracia oferece maiores oportunidades de liberdade e de autogestão, mas nem por isso deve ser considerada permissiva ou omissa. A accountability apresenta-se como princípio de organização política e administrativa tolerante, que promove a transparência pública e o controle social mediante a periódica prestação de resultados em ato público. A necessidade de saber o que, de fato, se aprende e não se aprende nas escolas brasileiras se torna urgente e perene neste contexto. A avaliação cognitiva se torna mais consistente e estável com recentes avanços em psicometria, viabilizando o monitoramento do perfil cognitivo em indivíduos, série-a-série, e em instituições ano-a-ano. A avaliação recorrente dos resultados alcançados nas escolas apresenta-se como antídoto natural às habituais oscilações entre intervencionismo grosseiro e omissão, permitindo que se distingua seriedade de leviandade na administração pública.