

A Mensuração Educacional - Problema geral

A análise dos instrumentos de medida do rendimento escolar, ora empregados em nosso meio educacional, revela diversos níveis de qualidade técnica. Ao lado de alguns poucos que realmente demonstram medir aquilo a que se propõem, existe, infelizmente, um número elevado de instrumentos que apresentam completa carência de requisitos técnicos. O problema da qualidade desses instrumentos é grave, sobretudo em virtude da influência que exercem no processo de aprendizagem.

Os instrumentos de medida, independentemente do seu aspecto formal, mas desde que bem construídos, representam um estímulo para o estudante e um desafio ao seu interesse e à sua curiosidade intelectual. É fato reconhecido que os bons instrumentos de medida exercem uma função direcional, pois orientam o examinando sobre o que estudar e, mais importante ainda, sobre como estudar. Entretanto, quando certos instrumentos de medida são analisados, observa-se que não orientam, mas sim conduzem o estudante a adotar comportamentos sem grande relevância educacional, ou seja, estimulam a aprendizagem do efêmero e do factual, e, assim, transformam-se num elemento de frustração para o estudante, o qual, contrariamente ao que se poderia acreditar, tem plena consciência de não estar sendo avaliado segundo as suas expectativas (EBEL, 1961).

Se há instrumentos de mensuração que não estimulam, não orientam e não avaliam o progresso do estudante, qual seria, então, a razão de ser dessa anomalia? Este estado de coisas não decorreria do tipo de instrumento construído e, particularmente, do tipo de questão elaborado?

* Pesquisador sênior do Departamento de Pesquisas Educacionais da Fundação Carlos Chagas.

A pergunta é ociosa e já foi respondida há mais de meio século através de pesquisas empíricas. Existem boas provas de dissertação, assim como existem bons testes objetivos. O problema não se concentra neste ou naquele tipo de prova, mas na ausência do domínio da tecnologia dos instrumentos de medida educacionais. Alguns responsáveis pela elaboração de provas e exames simplesmente improvisam, quase sempre de boa fé, e, assim, praticam enganos que geram numerosos problemas na área da mensuração educacional.

Implementação de Um Programa de Medidas Educacionais

O principal motivo para a implementação de um programa de medidas na escola é a coleta de informações sobre o estudante, para que o professor possa planejar o seu trabalho e avaliar o ensino que ministra. Este objetivo, entretanto, só é alcançado na medida em que o professor acredite na eficiência do ensino diferenciado, ou seja, se ele acreditar na necessidade de diferentes materiais, técnicas e objetivos adequados ao aluno. Se, ao contrário, achar possível ensinar as mesmas coisas, do mesmo modo, com o mesmo material e as mesmas técnicas a todos os alunos, independentemente da diversidade de suas capacidades e dos vários níveis de rendimento, então as informações que podem ser obtidas não terão maior significado para esse professor (MSU -GUIDANCE DEPARTMENT, 1959).

Além desse objetivo, e no mesmo nível de importância, um programa de medidas procura identificar a constelação de interesses de cada estudante, para proporcionar orientação educacional e vocacional. Um programa de medidas sempre está associado a um programa de orientação, pois todas as informações coletadas destinam-se a oferecer elementos para este fim.

Há, ainda, dois outros objetivos que todo programa procura realizar: um, na área curricular; outro, na esfera administrativa. Os planejadores de currículo, com base nas capacidades e nos níveis de desempenho dos estudantes, podem julgar quais as seqüências de experiências de aprendizagem mais adequadas aos diversos grupos de estudantes. Já os administradores, a partir das mesmas informações, podem avaliar a eficiência dos programas e determinar se a instituição está realmente realizando os seus objetivos.

O sucesso na implementação de um programa de medidas depende, preliminarmente, da consideração de diferentes elementos:

- fixação dos seus objetivos;
- decisão sobre os tipos de instrumentos a empregar;
- seleção de instrumentos para a mensuração de variáveis específicas.

Independentemente da sua amplitude, é necessário a organização de uma estrutura de base para garantia da execução e continuidade do programa, que nunca é obra exclusiva de um indivíduo, nem se concentra no trabalho de alguns poucos professores e orientadores. Ao contrário, um programa de medidas exige o trabalho cooperativo de todo o sistema institucional. Assim, impõe-se a criação de algumas equipes e serviços, a saber:

- equipe de especialistas em mensuração do rendimento escolar, de aptidões e de outras variáveis, segundo os objetivos do programa;
- equipe de orientadores com experiência na área vocacional;
- serviços de processamento de dados;
- serviços de análise estatística;
- serviços administrativos.

Um programa de medidas, por sua própria natureza, pressupõe alto grau de conscientização e proficiência de todos os elementos executivos. Essas características podem ser alcançadas através:

- do treinamento de professores e orientadores, para identificá-los com os objetivos do programa, do uso adequado dos instrumentos e do desenvolvimento da capacidade de interpretar os resultados;
- do estabelecimento de meios eficientes de comunicação entre especialistas em mensuração, professores e orientadores.

A implementação de um programa pressupõe, ainda: aplicação sistemática dos instrumentos durante o semestre ou ano letivo;

- construção e aplicação de testes similares em sua estrutura e áreas de conteúdo, em períodos letivos sucessivos;
- processamento rápido e preciso dos dados, se possível através de meios mecânicos;
- divulgação imediata dos dados e relatórios entre professores e orientadores;
- registro sistemático dos dados em fichários cumulativos individuais, constantemente atualizados e de fácil consulta.

Análise de Aspectos de Um Programa de Medidas

O controle da eficiência do programa de medidas é realizado por um programa paralelo de avaliação. Ainda que bem planejado e executado, nenhum programa dispensa uma avaliação que, inclusive, pode ser subjetiva e sem o rigor de um experimento empírico. Os objetivos do programa, o tipo da escola etc. são algumas das variáveis que podem determinar os critérios para esta avaliação; entretanto, independentemente dessas variáveis, que caracterizam o programa de medidas, o avaliador deve, obrigatoriamente, analisar determinados aspectos que são comuns a todos os programas e refletem a sua eficiência.

Os mais importantes destes aspectos (LEFEVRE, 1950; LENNON, 1954; TRAXLER, 1950) serão discutidos no presente tópico.

- Um programa procura coletar o maior número possível de informações, de forma a compor um perfil que caracterize o estudante em diversas dimensões e cuja análise possibilite a compreensão do seu desempenho em diferentes áreas. Assim sendo, um dos aspectos a avaliar é o da compreensibilidade dos instrumentos.
- Alguns programas incluem apenas grupos de voluntários, ou grupos especialmente selecionados ou, ainda, grupos com problemas específicos. Quando isto ocorre, há um desvirtuamento das verdadeiras funções do programa de medidas, pois as conclusões extraídas são válidas apenas para o grupo e não podem ser extrapoladas para todos os elementos da instituição, que deveriam integrar o programa. Quando ocorre essa limitação na concepção e organização do programa, há necessidade de promover a sua revisão, a fim de que os dados coletados permitam conclusões sobre a capacidade e o rendimento do grupo como um todo, ou seja, a universalidade da sua aplicação deve ser uma das características de qualquer programa de medidas.
- A aplicação dos instrumentos de medida depende muitas vezes de conveniências administrativas da escola; dessa forma, nem sempre ocorre quando o professor necessita de elementos para orientar a aprendizagem, planejar o ensino ou tomar medidas práticas que garantam o progresso do estudante. A regularidade na aplicação dos instrumentos deve ajustar-se ao processo de ensino e não constituir um fator restritivo da atividade do professor ou do orientador.
- Aspecto crítico dos programas de medidas é a comparabilidade dos instrumentos empregados. A similaridade na construção dos instrumentos nem sempre ocorre. Numa mesma instituição, com as mesmas áreas de conteúdo, num mesmo período letivo e para alunos do mesmo nível de escolaridade, são construídos instrumentos com diferentes especificações. Isso resulta num quadro confuso, que o mais hábil especialista não consegue interpretar; conseqüentemente, não são possíveis comparações entre os indivíduos do grupo.
- A fixação dos objetivos é o ponto inicial de qualquer planejamento. O problema, no caso dos programas de medidas, é bem mais complexo. Há necessidade de estabelecer uma concordância entre os objetivos da escola, do currículo e os do próprio programa. Os instrumentos desenvolvidos e empregados precisam estar em sintonia com todos esses objetivos. Exemplificando, suponhamos uma escola que objetive o ensino integrado. Não se justificaria que em Estudos Sociais se aplicassem testes individualizados de História e Geografia. Os resultados não teriam significado, por violarem os objetivos da escola (ensino integrado) e do currículo (Estudos Sociais).
- Existem três problemas que não devem ser ignorados no programa de avaliação: a maneira como os instrumentos são aplicados, o processo de correção e a qualificação dos responsáveis pela interpretação dos resultados.

As cautelas relativas à aplicação dos instrumentos podem parecer um exagero para os que não se dedicam à mensuração educacional. A prática demonstra, entretanto, que, apesar de todas as precauções, alguns resultados de testes, reconhecidamente válidos e fidedignos, carecem de valor em virtude da não-observância das normas de aplicação.

A correção dos instrumentos, ainda que por equipamento eletrônico, exige um trabalho suplementar, manual, de correção, em virtude de erros cometidos pelo estudante no

preenchimento das folhas de respostas ou dos cartões. Esses erros são acusados pelas máquinas de leitura óptica e registrados pelo computador ao realizar o teste de qualidade do exame que está sendo corrigido.

Os dados coletados são usados para fins de decisão educacional e vocacional. Antes de qualquer decisão, os dados precisam ser interpretados por pessoal qualificado e especialmente treinado para esse fim. A falta de qualificação técnica do usuário dos resultados de um programa de medidas pode gerar situações comprometedoras para o futuro educacional e vocacional do estudante.

Assim sendo, na avaliação de um programa de medidas é preciso verificar se:

- os aplicadores receberam treinamento adequado;
- as equipes de correção e recorrecção, organizadas com pessoal administrativo, foram devidamente alertadas e treinadas para a solução dos problemas dessa fase do programa;
- os responsáveis pela interpretação dos resultados possuem alto nível de qualificação técnica.

- Um programa inadequadamente conduzido tende a desenvolver estruturas burocráticas hipertrofiadas. Há coleta de informações, mas estas não atingem rapidamente os maiores interessados - professores e orientadores - e rapidamente se tornam obsoletas. Formam um verdadeiro arquivo morto.

É necessário que os resultados sejam rapidamente divulgados, juntamente com outros elementos essenciais para o trabalho dos professores e orientadores.

O registro desse material, em fichas cumulativas, ano após ano, é de grande importância porque, através da avaliação individual de cada aluno, em todas as áreas, e do acompanhamento dos seus ganhos educacionais, permite que os professores e orientadores disponham de elementos para julgar e encaminhar cada indivíduo segundo suas potencialidades e vocação profissional.

A avaliação de um programa de medidas precisa considerar, portanto, dentre outros aspectos, a maneira como os resultados são divulgados e a sistemática do seu registro.

- Os testes de rendimento escolar são importantes num programa de medidas, mas não devem constituir o único tipo de instrumento a ser empregado. É interessante também coletar o máximo possível de informações nas áreas cognitiva, afetiva e vocacional. Assim, além daqueles testes, é desejável o uso de outros instrumentos para a mensuração de inteligência, aptidões especiais, interesses e qualificações pessoais. Sem informações oriundas de várias fontes não é possível o estudo da inter-relação das variáveis que possibilitam a compreensão do estudante e a sua orientação.

A avaliação do programa deve verificar se houve equacionamento de diferentes variáveis e se estas estão sendo mensuradas por instrumentos hábeis.

- Qualquer que seja o tipo de programa - simples ou complexo, míni ou macro -, estará condenado ao fracasso se não contar com equipe capaz de usar satisfatoriamente as suas informações. Não se pretende que todos os professores e orientadores sejam especialistas em medidas, em testes e em Estatística, ainda que seus elementos básicos possam ser dominados por pessoas medianamente dotadas. A existência de especialista qualificado, que possa colaborar nesse treinamento e na organização e implementação do programa, é aspecto a ser analisado na avaliação de um programa de medidas.

As Informações Coletadas e Seus Objetivos

Os programas de medidas empregam diferentes tipos de instrumentos - testes de rendimento escolar, testes de aptidão geral, testes de aptidões específicas e diversos inventários - para coleta de informações, objetivando:

- determinar a aptidão escolar e o nível mental dos alunos para estabelecer comparações com estudantes de outras escolas do mesmo tipo ou de tipos diferentes, na mesma região ou em áreas diferentes. O conhecimento da capacidade mental média e do grau de homogeneidade, ou de heterogeneidade, da inteligência dos elementos do grupo é o ponto inicial para o planejamento e o desenvolvimento de currículos, assim como para o ajustamento dos métodos de ensino aos diferentes níveis de capacidade;
- comparar o rendimento de cada aluno com o rendimento previsto a partir do desempenho em teste de aptidão escolar. O grupo como um todo pode estar tendo um rendimento diferente daquele que se desejaria; individualmente, alguns elementos podem apresentar um desempenho superior ou inferior ao esperado e previsto com base nas medidas de aptidão.

Sobre o rendimento escolar é necessário observar os seguintes pontos:

- a mensuração do rendimento escolar nos primeiros anos da escola elementar é questionável, porque:
 - os escores dependem da capacidade de leitura;
 - os escores são influenciados pela memória mecânica;
 - uma pressão excessiva sobre as crianças pode modificar a sua atitude em relação à escola e afetar o seu sucesso posterior;
- os testes de rendimento e os de inteligência muitas vezes fornecem informações que podem ser usadas para predição do provável sucesso escolar; desse modo, o uso de testes de prognose apenas duplicará informações;
- alguns testes medem aspectos isolados da educação formal; outros enfatizam o uso de informações para demonstração de pensamento crítico, capacidade de resolver problemas e outros comportamentos complexos. O uso preferencial de um ou de outro instrumento refletirá a filosofia educacional da escola;
- a fidedignidade dos escores está associada à extensão do teste; assim, o emprego de testes curtos pode representar uma economia aparente à custa do grau de consistência dos escores;
 - proporcionar elementos para que o professor, com base no nível de capacidade e de rendimento dos alunos, possa planejar o nível e o tipo de ensino, estabelecendo diferentes níveis segundo a diversidade das capacitações. Em função destas, o processo pode ser acelerado ou retardado, conforme a capacidade de aprendizagem mais rápida ou mais lenta dos alunos;

- comparar o rendimento individual de cada aluno com a sua aptidão escolar, ou seja, verificar como cada aluno está progredindo em relação à capacidade demonstrada. O grupo, como referencial, algumas vezes leva a conclusões errôneas; assim, um aluno pode não estar progredindo em relação ao grupo, mas apresentar um bom rendimento quando o critério é o seu desempenho em testes de aptidão; inversamente, pode estar acompanhando o nível do grupo, mas ter um rendimento abaixo do seu potencial;
- diagnosticar deficiências do grupo e, particularmente, dos indivíduos, para, através de um trabalho conjunto de professores e orientadores, auxiliar o aluno a superar as suas deficiências;
- identificar e desenvolver aptidões para certas áreas, como ciências, línguas, artes etc.;
- desenvolver uma compreensão mais profunda dos interesses dos alunos. Através de observações e de entrevistas, da análise de leituras realizadas pelo aluno e de como este ocupa as suas horas de lazer, é possível identificar alguns interesses; contudo, em níveis mais avançados de escolaridade, são necessários inventários especiais para um conhecimento específico dos seus interesses;
- ajustar o indivíduo ao meio educacional. Os instrumentos usados num programa de medidas testes de aptidão, de rendimento escolar e de aptidões especiais - possibilitam a identificação de sintomas de desajustamento; entretanto, em certos casos, esse comportamento só é positivado através de inventários de personalidade. Levando-se em consideração a complexidade desses instrumentos e a necessidade de especialização para a interpretação de seus resultados, a escola somente deve empregá-los se dispuser de um psicólogo especialmente treinado nessa área; caso contrário, é recomendável o uso de métodos informais de avaliação da personalidade. A decisão da escola sobre o uso de inventários de personalidade precisa ser cautelosa. Antes de qualquer decisão, um exame cuidadoso das provas da validade do instrumento é necessário. Ainda que essas provas existam, é aconselhável obter a aquiescência dos pais, que devem ser esclarecidos sobre as funções e os objetivos de um inventário de personalidade, a fim de evitar problemas para a comunidade escolar. Finalmente, antes de qualquer decisão, a escola deve considerar as suas possibilidades de proporcionar tratamento ou de indicar recursos psicológicos ou psiquiátricos que possam ser úteis aos estudantes com problemas de ajustamento;
- orientar e aconselhar os alunos para que se sintam ajustados em suas aspirações vocacionais;
- desenvolver, a longo prazo, em cada estudante, a capacidade de autocompreensão, auto-avaliação e autodependência (TRAXLER, 1960; MSU - GUIDANCE DEPARTMENT, 1959).

Diferentes Modelos Para a Mensuração Educacional

Afirma-se algumas vezes, mas sem muita convicção, que estaria havendo, no momento, uma proliferação de testes, sobretudo nas escolas de 1° e 2° Graus. Acreditamos que não constitua malefício o uso de testes em qualquer dos níveis de escolaridade, desde que os instrumentos sejam tecnicamente idôneos, usados com propriedade nos casos indicados e os resultados interpretados por pessoa qualificada.

Analisando-se alguns instrumentos atualmente empregados, observa-se que o principal objetivo é coletar dados para a organização de uma rudimentar e discutível contabilidade do rendimento escolar. Além de não considerarem aspectos relevantes para a mensuração escolar - validade e fidedignidade -, esses instrumentos são inconseqüentes, porque não informam ao aluno sobre o

seu progresso e não possibilitam ao professor verificar a concretização dos objetivos educacionais. À deficiência na construção dos testes e provas associa-se o desconhecimento de técnicas estatísticas elementares; desse modo, o desempenho individual não é analisado em função do comportamento do grupo (LAIDLAW, 1965).

O estado atual das mensurações educacionais simplesmente demonstra que, na escola, por falta de recursos humanos com treinamento especializado, não estão sendo empregados diferentes modelos de mensuração para fins diversos, com evidentes prejuízos para o aluno, o professor e a educação.

Quais os diferentes modelos que poderiam ser utilizados? Testes referenciados a critério, testes referenciados a norma e testes combinados (critério e norma).

Os testes de critério poderiam informar até que ponto os objetivos de uma unidade (ou curso) foram realmente alcançados, possibilitando, conseqüentemente, ao professor uma orientação segura de unidades subseqüentes (ou cursos), sem defasagens no rendimento escolar, pois o teste de critério, através da fixação de um nível mínimo de competência (o critério), procura determinar o domínio pelo aluno de pré-requisitos; assim, o estudante, informado de seu sucesso ou insucesso, tem consciência do seu progresso, enquanto o professor, por sua vez, pode organizar programas de recuperação para os que não obtiveram êxito e, por intermédio de métodos e técnicas especiais, levar o aluno a superar suas deficiências de aprendizagem e a acompanhar, sem maiores problemas, o desenvolvimento de outras unidades ou cursos.

Há necessidade de outro modelo - o teste referenciado a norma - afim de verificar o desempenho relativo do indivíduo, comparando o seu rendimento com o do grupo. A função desse modelo não é a de verificar objetivos e determinar deficiências individuais, mas sim a de coletar informações que possam levar à tomada de decisões, como aprovação e orientação vocacional.

Ambos os testes - critério e norma - completam-se por suas informações; contudo, um terceiro modelo poderia ser construído para obter os mesmos elementos que estes dois proporcionam isoladamente. Ainda que mais complexo, o teste combinado é, na verdade, um teste de norma sobreposto a um teste de critério.

O teste combinado apresenta, inicialmente, um teste de critério, que inclui todos os objetivos a verificar. É o quadro de referência para a elaboração de um teste mais extenso. Os itens estabelecem o desempenho mínimo aceitável, por isso são fáceis, com um índice de 90%, porcentagem esperada de acertos.

A partir dos objetivos fixados, são elaborados novos itens para a verificação de diferentes níveis de desempenho além do mínimo aceitável. Os itens da parte de norma sofrem um aumento crescente de dificuldade, cuja amplitude deve variar entre 20% e 80%, a fim de discriminar os melhores. A correção do teste combinado é feita em dois momentos. Inicialmente, é corrigida a parte relativa ao teste de critério, não havendo escores, mas apenas sucesso ou insucesso em alcançar o mínimo aceitável. Os que não foram bem sucedidos são submetidos a diferentes formas de ensino de recuperação até que consigam atingir o critério. Os bem-sucedidos têm a segunda parte do teste (norma) corrigida e são atribuídos escores para fins vários.

Os instrumentos aplicados em nossa escola não se enquadram no modelo critério, ainda que estabeleçam um desempenho mínimo - "a média" -, porque o ensino não é orientado por objetivos, e os instrumentos não seguem a mesma orientação; por outro lado, quando o critério não é atingido, as possibilidades de recuperação são mínimas, talvez seja necessário um novo exame - "a recuperação" -, geralmente tão duvidoso quanto os exames anteriores, havendo a

ameaça de uma reprovação pura e simples, com a repetição de novo período letivo, cuja eficiência é discutível. Os mesmos testes também não podem ser considerados de norma, pois nem sempre possuem validade e quase nunca oferecem resultados fidedignos. O que são? O que medem? O que permitem avaliar? - são questões difíceis de elucidar no momento presente.

Introdução: Uma "Nova" Medida

A utilização de objetivos comportamentais, para fins de orientar o processo instrucional, o desenvolvimento da instrução seqüencial e individualizada, a utilização de diferentes formas de instrução programada no processo ensino-aprendizagem, o surgimento de novas metodologias para a avaliação de currículos e, especialmente, a elaboração de novas teorias sobre a aprendizagem contribuíram para o surgimento de um "novo" tipo de medida educacional - as medidas referenciadas a critério. Além disso, as discussões sobre os testes tradicionais e as críticas aos sistemas de notas empregados no contexto educacional concorreram para que o problema da avaliação de desempenhos, no âmbito escolar, fosse dimensionado segundo uma perspectiva diversa da habitual. O desenvolvimento da tecnologia educacional, por sua vez, criou expectativas quanto a um novo instrumental de mensuração que possibilitasse avaliar não apenas o desempenho de indivíduos submetidos à instrução, mas também a própria eficiência do processo instrucional.

GLASER (1963), ao advogar um novo modelo para a avaliação educacional, criou a expressão **testes referenciados a critério**. A sua preocupação centrava-se no problema da instrução individualizada e no da promoção do uso sistemático de objetivos comportamentais. Os instrumentos a serem empregados não poderiam ter as características dos tradicionalmente usados, que oferecem resultados em termos da posição relativa dos indivíduos no seu grupo. O importante para GLASER, no caso da instrução individualizada, era a situação do indivíduo em relação aos objetivos fixados, cuja aquisição traduziria o êxito do sistema instrucional. As medidas referenciadas a critério vieram ao encontro das expectativas geradas por essa abordagem do problema instrucional.

Outra questão relevante, igualmente equacionada pela abordagem teórica das medidas referenciadas a critério, diz respeito às diferenças individuais. Ainda que estas sejam freqüentemente acentuadas, os atuais sistemas instrucionais consideram os indivíduos indiferentemente, como grupos mais ou menos homogêneos, e os submetem a um único tratamento, no pressuposto de que todos os indivíduos, num mesmo período de tempo, e do mesmo modo, poderiam alcançar um mesmo nível de desempenho. Os sistemas educacionais, em geral, não consideram que diferentes indivíduos realizam o trabalho de aprendizagem com velocidades variadas, que a mesma aprendizagem exige diversificação de procedimentos para os diversos indivíduos e que, finalmente, os níveis alcançados são também variados. Alguns poucos alcançam apenas os padrões mínimos de suficiência; entretanto, a maioria, em princípio, poderia alcançar o nível de competência, se lhe fossem oferecidas condições adequadas. A presente teoria, evidentemente, não é nova, mas passou a ter grande repercussão com a divulgação dos trabalhos de CARROLL (1963) e de BLOOM (1968), que discutem, detalhadamente, o problema da aprendizagem segundo essa perspectiva.

As medidas referenciadas a critério ganharam, assim, novo alento com a difusão da teoria de que todos, ou quase todos, os indivíduos podem aprender e atingir um grau considerável de capacitação (BLOOM, 1968). A presente posição implicou, necessariamente, a rejeição da idéia de que a aprendizagem sofreria a influência de limitações inatas (AIRASIAN & MADAUS, 1974).

Por sua vez, a importante distinção estabelecida por SCRIVEN (1967) entre a avaliação formativa e somativa, na caracterização de uma metodologia para avaliação de currículo, concorreu, também, para acentuar o interesse por essa nova abertura no campo da avaliação instrucional.

Conceito de Medidas Referenciadas a Critério

Assim como o conceito de aprendizagem para o domínio não é inteiramente novo, também não constitui uma inovação o conceito de medida referenciada a critério. São novas, todavia, as abordagens ao problema. Quanto às medidas referenciadas a critério, poder-se-ia dizer que, de certa forma, os professores, nas suas atividades, costumam empregá-las, pois exigem um padrão mínimo para o sucesso. O exame de alguns aspectos ligados à avaliação e aos instrumentos de medida usados no dia-a-dia da instrução mostra, entretanto, que existem diferenças fundamentais entre a situação atual e a sugerida pelo novo conceito de medida.

Os instrumentos elaborados no decorrer da rotina instrucional nem sempre partem da especificação de objetivos estabelecidos anteriormente ao processo instrucional. A prática usual consiste na elaboração de questões relativas à área de conteúdo. Assim, o objetivo real do instrutor é verificar o que foi "lecionado". Este procedimento foge inteiramente à prática da avaliação por medidas de critério, a qual pressupõe um processo instrucional orientado por objetivos predeterminados. Outro aspecto a considerar refere-se ao grau de homogeneidade dos instrumentos. Ainda que os testes comuns sejam construídos a partir de objetivos comportamentais, afastam-se da abordagem das medidas referenciadas a critério, porque, habitualmente, são heterogêneos, isto é, verificam vários objetivos e/ou áreas de conteúdo, opondo-se, assim, às de critério, que são homogêneas.

O problema em questão é complexo e dificulta a interpretação dos desempenhos, pois o escore de um teste que mede dois ou mais objetivos pode levar a conclusões falsas, quando são estabelecidas comparações entre vários resultados. A título de exemplo, suponhamos um teste hipotético de vinte itens que meça quatro objetivos diferentes, cada um deles verificado por um conjunto de cinco itens. Admitamos que dois estudantes tenham obtido o escore quinze. O resultado não permite inferir que ambos os sujeitos tiveram desempenhos equivalentes, a não ser que as respostas sejam analisadas e se positive que, no caso, a diferença de cinco pontos correspondeu à não-concretização dos mesmos objetivos. O problema, numa situação de medidas referenciadas a critério, estaria superado, em virtude da homogeneidade desse tipo de instrumento.

Suponhamos, também, que diversos instrumentos de medida versem sobre as mesmas áreas de conteúdo. O significado de escores iguais não é pacífico, pois os instrutores diferem, às vezes profundamente, no que diz respeito à definição e à ênfase relativa que atribuem a cada objetivo. Acresce, ainda, que os instrutores costumam usar diferentes tipos de questões e, em muitos casos, apresentam variações no processo de correção, mesmo quando as provas são do chamado tipo objetivo. Notas ou escores iguais não significam, necessariamente, desempenhos iguais, aspecto este que as medidas de critério procuram solucionar.

As provas usadas na rotina instrucional visam coletar elementos para uma avaliação somativa, após uma seqüência instrucional, e, geralmente, limitam-se a proporcionar uma "nota". Ao contrário, as medidas referenciadas a critério são aplicadas em diferentes momentos do processo instrucional, para fins diversos das provas comuns. Quando usadas antes do processo instrucional, procuram verificar se o estudante possui os pré-requisitos para acompanhar a

seqüência instrucional; durante o processo, para positivar a ocorrência de problemas de aprendizagem e sugerir medidas corretivas de possíveis desvios, e, ao término do processo, para verificar se o desempenho correspondeu ao critério padrão estabelecido.

As medidas referenciadas a critério desempenham, portanto, papel nuclear na avaliação formativa da aprendizagem. As suas informações interessam ao estudante porque, ao especificarem erros de aprendizagem que precisam ser corrigidos, são uma motivação para a aprendizagem bem-sucedida. As mesmas informações são também significativas para os professores, porque permitem reestruturar estratégias deficientes e prejudiciais ao êxito instrucional.

O contraste entre medidas de critério e de norma pode ser estabelecido com base na interpretação habitualmente dada aos escores de indivíduos submetidos ao processo instrucional. Os escores oferecem duas categorias de informações:

- a posição relativa dos elementos no grupo, estabelecida em função do desempenho dos demais indivíduos na mesma escala;
- se os indivíduos, com os desempenhos apresentados, atingiram determinado critério preestabelecido.

Quando a interpretação de um escore é feita através da comparação com os escores dos outros indivíduos do grupo, a medida é referenciada a norma. O grupo constitui a norma que possibilita interpretar o significado do escore. Nesse caso, o escore informa se o treinando é mais ou menos capaz do que outros elementos do seu grupo; entretanto, nada indica relativamente à capacidade de cada indivíduo nas tarefas exigidas. Assim, no caso destas medidas, o desempenho, expresso por um escore, somente adquire significado quando uma relação entre os desempenhos de vários elementos é estabelecida. Duas indagações, importantes para o processo de avaliação, deixam de ser respondidas: quais os objetivos medidos pelo instrumento? e qual o nível de desempenho do grupo?

Se, ao contrário, os escores dos indivíduos forem relacionados a um critério de desempenho, a medida é referenciada a critério e procura refletir o status do indivíduo em relação a um determinado padrão de desempenho (POPHAM & HUSEK, 1969). Os indivíduos, no presente caso, são comparados a um padrão indicativo da concretização de um objetivo instrucional predeterminado (critério) e, assim, os seus escores refletem a sua maior ou menor competência, independentemente do referencial formado pelo grupo a que pertencem.

As técnicas de avaliação, conseqüentemente, variam nos dois casos. Se a interpretação baseia-se no desempenho de grupos, os instrumentos são construídos de forma a que proporcionem escores com grande amplitude de variação, o que é conseguido pela eliminação de itens fáceis ou difíceis, e pela utilização de questões de dificuldade mediana, que, em princípio, tendem a maximizar diferenças, o que permite uma melhor classificação dos indivíduos (GRONLUND, 1976). Se a interpretação visa a comportamentos específicos, certos critérios importantes para a seleção de questões no contexto das medidas referenciadas a norma (dificuldade e poder discriminativo) não prevalecem. A seleção dos itens é, então, realizada com base na validade de conteúdo, isto é, refletem os objetivos que devem ser medidos. Não há interesse em eliminar questões fáceis ou introduzir questões que promovam a variabilidade dos escores, pois no caso das medidas referenciadas a critério, conforme GLASER & NITKO (1971), os instrumentos são deliberadamente construídos para proporcionar medidas diretamente interpretáveis em termos de um domínio específico de tarefas instrucionais relevantes.

A distinção básica entre as medidas ora em discussão não está, todavia, no seu aspecto formal, mas sim nas informações que proporcionam. Ambas são empregadas para a tornada de decisões,

mas em contextos diferentes. No caso de medidas referenciadas a norma, a situação em que são empregadas apresenta uma certa seletividade (POPHAM & HUSEK, 1969), impondo-se, portanto, uma comparação entre as posições relativas dos diferentes sujeitos. Quando, ao contrário, se deseja estabelecer determinada competência de um indivíduo, as medidas referenciadas a critério são as mais indicadas.

A diferença fundamental entre as duas variedades de medidas está no tipo de informação que esses instrumentos proporcionam ao avaliador, a fim de que tome decisões (COX, 1972). Suponhamos, para exemplificar, que após uma seqüência instrucional foi aplicado um teste. Os resultados podem ser interpretados como referenciados a norma ou a critério. Se o conjunto dos estudantes for considerado um grupo normativo, aqueles que se situam acima de um desvio padrão da média receberiam conceito A. Aos demais seriam atribuídos outros conceitos conforme as suas posições, em termos de desvios padrão e em relação à média do grupo. A interpretação dos resultados, no presente caso, está sendo feita com referência a uma norma, pois o desempenho de cada indivíduo foi comparado ao dos outros elementos do grupo.

Se, no exemplo apresentado, fosse estabelecido que apenas os estudantes que responderam corretamente a 90% dos itens do instrumento prosseguiriam na seqüência instrucional, o enfoque seria de medidas referenciadas a critério. O padrão de desempenho mínimo aceitável é o critério, e o fato de o estudante ter ou não conseguido alcançá-lo permitirá ao professor orientar o processo instrucional e, portanto, a aprendizagem do aluno.

Os escores de um estudante caracterizam-se por duas dimensões: a primeira, a descritiva, procura descrever a magnitude do traço medido, enquanto a segunda, a avaliativa, tenta interpretar essa descrição, a fim de possibilitar uma tomada de decisão. Suponhamos, para análise, que um instrumento com vinte itens seja aplicado e os resultados, expressos em percentis. Quando os escores brutos são utilizados, apenas apresentam a dimensão descritiva e nada mais; entretanto, se associados à norma percentil, os novos escores, além de descritivos, oferecem, também, uma dimensão avaliativa. O desempenho pode ser considerado "bom" ou "ruim", conforme o percentil associado ao escore bruto. Esta é uma situação típica das medidas referenciadas a norma, que, entretanto, não é satisfatória, pois nada esclarece quanto ao nível real de capacitação.

As duas dimensões aplicam-se, igualmente, aos resultados das medidas de critério. Assim, supondo-se que se aplicou um teste referenciado a critério, o total de respostas corretas também apresenta uma dimensão descritiva, mas a sua componente avaliada não se associa ao desempenho geral do grupo e sim ao nível de desempenho considerado crítico. Se, por hipótese, o nível crítico (critério) foi fixado no escore quinze, os estudantes que o superaram são considerados "sucessos", "aptos", "capazes", ou "aprovados", e os que obtiveram escores inferiores são tidos como "insucessos", "inaptos", "incapazes", ou "reprovados". A interpretação da dimensão avaliativa, neste caso, é, portanto, absoluta e axiomática (AIRASIAN & MADAUS, 1974), havendo "sucesso" sempre que o estudante for capaz de apresentar determinado produto ou manifestar certo comportamento, segundo os objetivos do processo instrucional. Não existe preocupação em estabelecer "o quanto" é capaz em relação aos outros elementos do grupo.

As medidas ora apresentadas como alternativa para alguns problemas da avaliação instrucional não excluem a aplicação das medidas referenciadas a norma. Ambas são igualmente legítimas para fins de avaliação educacional, pois em algumas circunstâncias deseja-se discriminar os indivíduos segundo o nível de rendimento, enquanto em outras pretende-se verificar a concretização de um objetivo isolado ou de um conjunto de objetivos comportamentais. Ainda que

alguns educadores ponham em dúvida a eficiência das medidas referenciadas a critério (EBEL, 1972), por adotarem posição contrária aos novos modelos instrucionais e discordarem dos enfoques baseados em objetivos comportamentais, estas possuem potencialidades a explorar em benefício do processo instrucional e da avaliação educacional.

Medidas Referenciadas a Critério: Alguns Problemas

A análise dos contrastes entre os dois tipos de medidas põe em evidência alguns problemas básicos que precisam ser considerados, sobretudo quando se constata que as atuais práticas de avaliação educacional se acham impregnadas de elementos da teoria clássica dos testes, que, freqüentemente, são extrapolados para situações de medidas referenciadas a critério, apesar de estas últimas se basearem em pressupostos teóricos diferentes.

Ainda que não existam diferenças formais entre os itens utilizados nos dois instrumentos de medida ora discutidos, os comportamentos dos construtores de testes variam nas duas situações. Os testes referenciados a norma são construídos visando à coleta de informações sobre diferenças entre indivíduos; assim, procuram obter o máximo de discriminação. Os construtores de testes, portanto, esperam que os escores apresentem a maior variabilidade possível. O conceito de variabilidade dos indivíduos é, assim, pressuposto básico nessas medidas, o que está coerente com o seu principal objetivo: discriminar para estabelecer a posição relativa de cada indivíduo no seu grupo.

A abordagem é diferente nos testes referenciados a critério: os itens devem representar comportamentos definidos pelo critério. A facilidade ou dificuldade do item não é requisito básico para o seu julgamento; o que importa realmente é o fato de traduzir situações relevantes, que possibilitem demonstrar a aquisição dos comportamentos definidos pelos objetivos instrucionais (POPHAM & HUSEK, 1969).

Um item respondido por todos ou quase todos os sujeitos seria considerado muito fácil, e eliminado dos testes referenciados a norma; entretanto, no contexto dos testes referenciados a critério, o item seria incorporado ao instrumento, desde que adequado a verificar a concretização dos objetivos terminais, independentemente de todos o responderem corretamente, pois não existe a preocupação de promover a variabilidade dos indivíduos, pelo menos nos termos propostos pela teoria clássica dos testes. A variabilidade dos escores pode existir, e isso efetivamente ocorre nos resultados dos testes referenciados a critério, mas não constitui condição essencial para a caracterização desse tipo de instrumento. O problema, muitas vezes objeto de discussões (WOODSON, 1974, 1974a; MILMAN & POPHAM, 1974), deve ser examinado numa perspectiva diferente da que oferece a teoria das medidas referenciadas a norma.

A questão da discriminação, e, conseqüentemente, da variabilidade nas medidas referenciadas a critério, é equacionada em termos dos que atingiram, ou não, o critério estabelecido. Há uma situação ideal e, por isso mesmo, nem sempre concretizável: todos erram no pré-teste, todos acertam no pós-teste. Esta diferença representaria a variabilidade máxima nesses testes. Os fatos, na prática, apresentam-se de maneira diferente. Sempre há alguma variação, ainda que pequena, entre os que atingiram ou não o critério, após uma seqüência instrucional. O problema, portanto, não é uma questão de tudo ou nada.

Implementação de Um Programa de Medidas

A avaliação educacional objetiva, fundamentalmente, aprimorar o processo instrucional. O emprego de medidas referenciadas a critério, especialmente no contexto da avaliação formativa (BLOOM, 1968; GRONLUND, 1976), visa alcançar objetivos específicos que podem ser sintetizados da seguinte forma:

- corrigir deficiências da aprendizagem, ajustando, quando for o caso, o nível da instrução ao grupo, ou modificando estratégias instrucionais ou, ainda, introduzindo elementos necessários à facilitação da aprendizagem;
- orientar indivíduos ou grupos que apresentem deficiências de aprendizagem;
- promover a motivação de indivíduos através da indicação dos objetivos alcançados e do progresso de cada estudante, por intermédio de informações imediatas e específicas;
- aumentar a retenção da aprendizagem, tendo em vista que os instrumentos, nas medidas referenciadas a critério, procuram enfocar elementos importantes e, ao que tudo sugere, quanto mais complexo for o comportamento, maior a tendência à retenção;
- facilitar a transferência da aprendizagem, por intermédio de sua aplicação a novas experiências, similares às desenvolvidas no contexto instrucional.

A implementação de um programa de medidas referenciadas a critério, para atingir estes objetivos, pressupõe:

- definição, antes do início do processo instrucional, dos objetivos terminais indicadores dos desempenhos, habilidades e produtos que serão concretizados através do processo instrucional;
- decisão sobre os padrões - um padrão para cada objetivo e para cada conjunto de comportamentos critério - que serão utilizados para julgamento dos desempenhos ou produtos indicados nos objetivos. Ainda que o padrão seja estabelecido de maneira arbitrária, sua determinação baseia-se no julgamento lógico de especialistas, que analisam os objetivos para os quais o instrumento foi construído. Um nível crítico alto ou baixo depende, naturalmente, das conseqüências que as decisões tomadas, a partir do padrão, possam ter no processo instrucional. Supondo-se, com referência a um instrumento hipotético, que o seu objetivo seja o de verificar um módulo fundamental para a seqüência da aprendizagem, a fixação de um nível crítico baixo, por exemplo o escore dez, numa escala de zero a trinta, poderia afetar, seriamente, o desenvolvimento instrucional, em virtude das possíveis defasagens de muitos indivíduos em relação a comportamentos relevantes para a continuidade do processo ensino-aprendizagem. Assim, a fixação de padrões pressupõe a consideração das possíveis implicações que um determinado padrão ou critério possa ter na aprendizagem dos indivíduos. Alguns educadores sugerem ser possível um desempenho perfeito, correspondente a 100% dos objetivos; entretanto, uma posição mais realista será a de admitir-se um desempenho de 80 a 90% de eficiência;
- planejamento de situações que permitam ao estudante demonstrar habilidades, comportamentos ou qualquer outro produto resultante do tratamento instrucional. Ou seja, planejar instrumentos necessários ao trabalho de avaliação, os quais, conforme os objetivos a verificar, podem ser, ou não, testes de papel e lápis. O importante, no caso, é que todos os itens ou tarefas representem de maneira precisa o objetivo ou os objetivos definidos pelo critério, a fim de que inferências possam ser feitas a partir dos resultados obtidos.

As medidas referenciadas a critério são um caminho a explorar, sobretudo quando a preocupação maior, no processo instrucional, é proporcionar uma aprendizagem eficiente. Estas medidas ainda não são aplicadas no dia-a-dia da escola brasileira; entretanto, considerando-se que, no momento atual, o problema de recuperação de estudantes deficientes é crítico, especialmente na escola de 1º Grau, o seu emprego, com vistas a uma avaliação formativa, é uma alternativa que merece ser considerada e discutida. Alunos e professores lucrariam com a adoção de um sistema de avaliação com base nessas medidas, principalmente os primeiros, porque poderiam ser adequadamente orientados ao longo do processo instrucional, evitando-se, assim, o traumatizante fracasso ao final de uma seqüência que deve ser sobretudo formativa.

Referências Bibliográficas

- AIRASIAN, P.W. & MADAUS, G.F. Criterion referenced testing the classroom. In: TYLER, R.W. & WOLF, R.M. *Crucial issues in testing*. Berkeley, Califórnia, MsCutchan Publishing Corporation. 1974.
- BLOOM, B.S. Learning for Mastery. *Evaluation Comment*, Los Angeles, Center for the Study of Evaluation of Instructional Programmas, 1(?), 1980.
- EBEL, R.L. Improving the competence of teachers in educational measurement. *The Clearing House*, 1961.
- . *Essentials of educational measurement* Englewood Cliffs, Prentice-Hall, Inc., 1972.
- GLASER, R. & NITKO, J. Measurement in learning and instruction. In: THORNDIKE, R.L. *Educational Measurement*, Washington, D.C., American Council of Education, 1971.
- GROUNLUND, N.E. *Measurement and evaluation in teaching*. 3.ed. Nova Iorque. McMillan Company, 1976.
- LIDLAW, W.J. Teacher-made test; models to serve specific needs. *The Clearing House*, February, 1965.
- LEFEVER, D.W. The teacher's role in evaluating pupil achievement. *Education*. Indianapolis, Ind. The Bobbs Merrill Company, Inc., 1950.
- LENNON, R.T. Testing - bond or barrier between pupil and teacher? *Education*. Indianapolis, Ind. The Bobbs Merrill Company, Inc., 1954.
- LEST'S look at our testing program. Michigan State University. Guidance Department, Bureau of Educational Research, East Lansing, Mich., 1959.
- MILLMAN, J. & POPHAM, W.J. The issue of item and test variance for criterion-referenced tests a clarification. *Journal of Educational Measurement*, 1969.
- POPHAM, W.J. & HUSEK, T.R. Implications of criterion-referenced measurement. *Journal of Educational Measurement*, 1969.
- SCRIVEN, M. The methodology of evaluation. *Perspectives of Curriculum Evaluation*, Chicago, Rand MsNally, (1), 1967.

- TRAXLER, A.E. Fifteen criteria of a testing program. *The Clearing House*, (25), 1950.
- . Use of results of large-scale testing programs in instruction and guidance. *Journal of Educational Research*, (54) 1960.
- VIANNA, H.M. *Introdução à avaliação educacional*. São Paulo, JBRASA, 1989.
- . A perspectiva das medidas referenciadas a critério. *Educação e Seleção*, São Paulo, Fundação Carlos Chagas, (2): 5-14, dez., 1980.
- . Programa de medidas para fins de ensino e orientação. *Educação e Seleção*, São Paulo. Fundação Carlos Chagas, (11): 87-95, jan./jun., 1985.
- . Qualificação técnica e construção de instrumentos de medida. *Educação e Seleção*, São Paulo. Fundação Carlos Chagas, (10): 43-49, jul./dez., 1984.