

A PERSPECTIVA DAS MEDIDAS REFERENCIADAS A CRITÉRIO

Heraldo Marelim Vianna*

1.0 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 empregar 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, seria a *situação do indivíduo em relação aos objetivos fixados, cuja aquisição traduziria o êxito do sistema instrucional*. As medidas referenciadas a critério vieram de encontro às expectativas geradas por essa abordagem do problema instrucional.

* Do Departamento de Seleção de Recursos Humanos da Fundação Carlos Chagas.

Outra questão relevante, igualmente equacionada pela abordagem teórica das medidas referenciadas a critério, diz respeito às diferenças individuais. Ainda que sejam frequentemente 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 lhes 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, na rejeição da idéia de que a aprendizagem sofreria a influência de limitações inatas (Airasian e Madaus, 1974). Por sua vez, a importante distinção estabelecida por Scriven (1967) entre 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.

2.0 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 20 itens que meça quatro objetivos diferentes, cada um deles verificado por um conjunto de cinco itens. Admitamos que dois estudantes tenham obtido o escore 15. O resultado não permite inferir que ambos os sujeitos tiveram desempenhos equivalentes, a não ser que as respostas sejam analisadas e positivo-se 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 a 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 pôs-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.

3.0 MEDIDAS REFERENCIADAS A CRITÉRIO E A NORMAS: COMPARAÇÃO

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

1º) a posição relativa dos elementos no grupo, estabelecida em função do desempenho dos demais indivíduos na mesma escala; e

2º) 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 e 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 a utilização de questões de dificuldade mediana, que, em princípio, tendem a maximizar diferenças, o que permite a 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 (dificulda-

de e poder discriminativo) não prevalecem. A seleção dos itens é, então, realizada com base na validade de conteúdo; isto é, se 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 e 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 tomada de decisões, mas em contextos diferentes. No caso das medidas referenciadas a normas, a situação em que são empregadas apresenta uma certa seletividade (Popham e 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 *normas* ou a *critérios*. 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 que 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 como “bom” ou “ruim”, conforme o percentil associado ao escore bruto. Esta é uma situação típica das medidas referenciadas a normas, 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 avaliativa 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 15, 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 e 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 normas. Ambas são igualmente legítimas para fins de avaliação educacional, pois, em algumas circunstâncias, deseja-se

discriminar os indivíduos segundo o nível de rendimento; 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, 1970), 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.

4.0 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, frequentemente, 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 a 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 e Husek, 1969).

Um item respondido por todos ou quase todos os sujeitos seria considerado muito fácil, e eliminado dos testes referenciados a normas; 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; Millman e Popham, 1974), deve ser examinado numa perspectiva diferente da que oferece a teoria das medidas referenciadas a normas.

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.

5.0 VERIFICAÇÃO DA SENSIBILIDADE AO PROCESSO INSTRUCIONAL

A preocupação maior das medidas referenciadas a critério centraliza-se no *status* dos indivíduos submetidos ao processo instrucional; entretanto, quando os resultados dessas medidas são avaliados em termos de “sucesso e insucesso”, “aprovado ou reprovado”, ou “atin-

giu o critério... e não atingiu o critério...” é possível que algumas diferenças individuais, que seriam de interesse para a avaliação da eficiência instrucional, sejam negligenciadas. A diferença existente entre o comportamento no pré e no pós-teste, ao término de um período de instrução, pode oferecer informações úteis para o sistema instrucional.

Os testes referenciados a critério são formalmente organizados com um pequeno conjunto de itens, geralmente de múltipla escolha. O exame das respostas erradas, por meio da análise de cada item, nas duas situações, pré e pós-teste, permite estabelecer se houve ou não sensibilidade ao processo instrucional. A Tabela 5.1 mostra o desempenho de 20 alunos em cinco itens, antes e após uma seqüência instrucional.

Tabela 5.1 Respostas a cinco itens, antes e após a instrução; sua sensibilidade e classificação.

Itens	PRÉ-TESTE		PÓS-TESTE		Sensibilidade	Classificação
	Certo	Errado	Certo	Errado		
1	—	20	20	—	1.00	ideal
2	20	—	20	—	0.00	fácil
3	—	20	—	20	0.00	difícil
4	20	—	—	20	-1.00	deficiente
5	2	18	19	1	0.85	eficiente

Examinando-se a diferença dos comportamentos, pode-se definir sensibilidade à instrução como

$$S = \frac{C_{\text{pós}} - C_{\text{pré}}}{N}$$

em que

- S = sensibilidade à instrução
- $C_{\text{pós}}$ = número de acertos no pós-teste
- $C_{\text{pré}}$ = número de acertos no pré-teste
- N = número de alunos que responderam o item no pré e no pós-teste.

O item 1 mostrou total sensibilidade à instrução e representa uma situação ideal; pois todos erraram antes da instrução e acertaram no pós-teste, o que reflete, sem dúvida, uma grande eficiência do processo instrucional. Os itens 2 e 3, ao contrário, não apresentam qualquer sensibilidade; pois, no caso do item 2, os alunos já possuíam os conhecimentos e/ou comportamentos, enquanto que no item 3, extremamente difícil, a instrução não concorreu para qualquer mudança. O item 4 revelou possível distorção do processo instrucional, tendo em vista o inesperado da situação no pós-teste. O exame da matriz de respostas possivelmente aconselhará mudanças na estratégia instrucional. O último item apresenta uma situação freqüente — a maioria consegue um desempenho correto após a instrução. Assim, quanto maior o coeficiente de sensibilidade, que pode variar de -1,00 a +1,00, maior, igualmente, a influência do processo instrucional na aprendizagem.

Suponhamos que, após duas aplicações de um teste, seja construída uma matriz 3 x 3 das omissões, respostas corretas e erradas de um item (Tabela 5.2). A matriz que se obtém apresenta um sumário da sensibilidade à instrução, ou seja, mostra a existência ou não de ganhos educacionais.

Tabela 5.2 Omissões, respostas corretas e erradas de um item de TRC, no pré-teste e no pós-teste.

Respostas	PÓS-TESTE			
	Omissões	Corretas	Erradas	Total
Omissões	7	9	17	33
Corretas	36	323	515	874
Erradas	27	88	298	413
Total	70	420	830	1.320

O número de indivíduos que acertou o item, antes do processo instrucional, foi 420 (32%). Este número, no pós-teste, aumentou para 874 (66%), havendo, pois, uma diferença de 34% entre os desempenhos, antes e após a instrução. As omissões, inicialmente, chegaram a 70 (5%) e no pós-teste modificaram-se para 33 (3%). Aparentemente, uma das conseqüências da instrução está na redução dos que, por razões várias, inclusive falta de motivação, omitiram o item.

Analisando-se a Tabela 5.2 observa-se que 454 sujeitos, após a instrução, mudaram o seu comportamento e, juntamente com os 420 que acertaram o item no pré-teste, responderam corretamente à situação proposta. Esta diferença entre o pós e o pré-teste constitui o resultado da instrução. Entretanto, há um problema a considerar na presente matriz, porquanto 88 indivíduos, que acertaram o item antes da instrução, mudaram a sua resposta no pós-teste, errando-a, o que também traduz um resultado da instrução, que deve ser pesquisado.

Outros elementos sobre a maior ou menor sensibilidade dos estudantes podem ser coletados se as respostas ao item forem apresentadas sob a forma de uma matriz 5 x 5 (Tabela 5.3).

Tabela 5.3 Omissões e respostas às alternativas de um item de TRC, no pré-teste e no pós-teste.

Pós \ Pré	O	A	B	C*	D	Total	%
	O	7	5	5	9	7	33
A	4	17	16	12	18	67	5
B	5	20	15	21	20	81	6
C*	36	150	166	323	199	874	66
D	18	46	54	55	92	265	20
Total	70	238	256	420	336	1.320	100
%	5	18	19	32	26	100	

C* — alternativa correta.

A análise das distribuições marginais ressalta a insensibilidade da alternativa D à instrução. Quando da aplicação do instrumento antes da instrução, 26% dos indivíduos responderam D e após a instrução 20% mantiveram a mesma resposta, enquanto que, para as demais alternativas,

houve sensível decréscimo após o módulo instrucional, o que permite acreditar que a alternativa D contenha elementos que não foram afetados pelo processo.

Se o exame da matriz se limitar apenas às respostas erradas, a insensibilidade da alternativa D à instrução é ainda mais evidenciada (Tabela 5.4). A nova matriz confirma a suspeição anterior e pode servir de subsídio para a revisão do processo instrucional, a fim de eliminar esse fator de perturbação.

Tabela 5.4 Respostas erradas às alternativas de um item de TRC, no pré e no pós-teste.

	O	A	B	C*	D	Total
Pré-teste	70	238	256	—	336	900
%	8	26	28	—	37	99
Pós-teste	33	67	81	—	265	446
%	7	15	18	—	59	99

C* — alternativa correta.

A porcentagem das respostas erradas mostra que as alternativas A e B foram sensíveis à instrução, o mesmo não acontecendo com a alternativa D, que apresentou um aumento relativo de 37 para 59%, o que mais uma vez confirma a sua insensibilidade à instrução. Outras conclusões sobre a maior ou menor eficácia do processo instrucional são possíveis quando se consideram as proporções das respostas numa matriz 5 x 5. A Tabela 5.5 reproduz os dados da Tabela 5.3 em termos de proporções.

Tabela 5.5 Proporções das omissões e respostas às alternativas de um item de TRC, no pré e no pós-teste.

Pós	Pré	O	A	B	C*	D
	O	0,10	0,02	0,02	0,02	0,02
A	0,06	0,07	0,06	0,03	0,05	
B	0,07	0,08	0,06	0,05	0,06	
C*	0,51	0,63	0,65	0,77	0,59	
D	0,26	0,19	0,21	0,13	0,27	

C* — alternativa correta.

A Tabela 5.5 permite verificar que os que omitiram o item no pré-teste têm menor probabilidade de êxito no pós-teste (0,51), quando comparados aos que deram alguma resposta, mesmo errada, no pré-teste; o que indica, provavelmente, que o grupo dos que omitem é menos sensível à instrução e, portanto, necessita de maiores cuidados durante o processo instrucional.

Observa-se, também, que a instrução tende a mudar o comportamento do indivíduo, qualquer que tenha sido a sua resposta inicial, conforme demonstram as proporções dos que responderam errado no pré-teste e depois acertaram no pós-teste (A - 0,63; B - 0,65 e D - 0,59).

A mesma Tabela 5.5 põe em destaque a alternativa D, ressaltando a sua resistência à instrução. Os que omitiram no pré-teste e depois responderam D no pós-teste, assim como os que

responderam D antes e depois do processo instrucional, constituem elevadas proporções, o que gera a suspeição de possivelmente existir nessa alternativa algum elemento de perturbação. Observa-se, ainda, que essa tendência é aproximadamente igual para os diversos subgrupos.

Outro ponto a destacar é o da existência de uma tendência à repetição do erro. A resistência ao fator instrucional, quando o processo é bem estruturado, costuma ser moderada, quase insignificante, mas não deve ser desprezada no processo de avaliação formativa. O presente fato ressalta da observação dos padrões de resposta O-O (0,10), A-A (0,07) e B-B (0,06). A existência destes padrões pode ser eliminada pelo instrutor por intermédio de mudanças em sua estratégia instrucional.

A análise das respostas ao item de instrumentos de medidas, aplicado antes e depois do processo instrucional, oferece possibilidades para estabelecer quais os elementos mais ou menos sensíveis ao processo instrucional. O exame das matrizes de respostas, por outro lado, fornece um "feedback" para o instrutor, pois revela as conseqüências da orientação que imprimiu ao treinamento e possibilita, se for o caso, a mudança de suas estratégias instrucionais.

6.0 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 a alcançar objetivos específicos que podem ser sintetizados da seguinte forma:

- 1) 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;
- 2) orientar indivíduos ou grupos que apresentem deficiências de aprendizagem;
- 3) promover a motivação de indivíduos através da indicação dos objetivos alcançados e do progresso de cada estudante, por intermédio de informações imediatas e específicas;
- 4) aumentar a retenção da aprendizagem, tendo em vista que os instrumentos, nas medidas referentes a critério, procuram enfatizar elementos importantes e, ao que tudo sugere, quanto mais complexo for o comportamento, maior a tendência à retenção; e
- 5) 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 a estes objetivos, pressupõe:

19) *Definição, antes do início do processo instrucional, dos objetivos terminais indicadores dos desempenhos, habilidades e produtos que serão concretizados através do processo instrucional.*

20) *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 estabelecido de maneira arbitrária, a determinação do padrão 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 10, numa escala de 0 a 30, 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 pos-

sa 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.

39) *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, sobretudo 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. e MADAUS, G.F. (1974) – Criterion referenced testing in the classroom – in TYLER, R.W. e WOLF, R.M. *Crucial Issues in Testing*. MrCutchan Publishing Corporation. Berkeley. California.
- BLOOM, B.S. (1980) – Learning for Mastery. *Evaluation Comment*, 1., nº 2. Center for the Study of Evaluation of Instructional Programs. Los Angeles. California.
- GLASER, R. e NITKO, J. (1971) – Measurement in learning and instruction – in THORNDIKE, R.L. – *Educational Measurement*. Washington D.C. – American Council on Education.
- GRONLUND, N.E. (1976) – *Measurement and evaluation in teaching*. Third Edition. Nova Iorque. McMillan Company.
- MILLMAN, J. e POPHAM, W.J. (1974) – The issue of item and test variance for criterion-referenced tests – a clarification. *Journal of Educational Measurement*, 11, 2.
- POPHAM, W.J. e HUSEK, T.R. (1969) – Implications of criterion-referenced measurement. *Journal of Educational Measurement*, 6, 1.
- SCRIVEN, M. (1967) – The Methodology of Evaluation – in *Perspectives of Curriculum Evaluation*. AERA Monograph, nº 1. Chicago. Rand McNally.