

QUALIFICAÇÃO TÉCNICA E CONSTRUÇÃO DE INSTRUMENTOS DE MEDIDA EDUCACIONAL

Heraldo Marelim Vianna*

1.0. 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 elaborada? A pergunta é ociosa e já foi respondida há mais de meio século através de pesquisas empíricas. Existem boas provas de dissertação, assim como há bons testes objetivos. O problema não se concentra neste ou naquele tipo de prova, mas na ausência do domínio da tec-

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

nologia dos instrumentos de medida educacional. 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.

2.0. OS INSTRUMENTOS E AS MEDIDAS EDUCACIONAIS – PROBLEMAS ESPECÍFICOS

O exame de alguns problemas relacionados com o processo de mensuração demonstra a falta de formação técnica de muitos construtores de instrumentos de medida. Apenas os problemas mais significativos serão discutidos a seguir.

2.1. – A maioria dos julgamentos sobre o rendimento educacional é subjetiva, ainda que muitos avaliadores acreditem possuir padrões absolutos de julgamento.

A carência de fidedignidade dos julgamentos subjetivos acha-se demonstrada por copiosa literatura de pesquisas em educação. A solução para o problema estaria na realização de um julgamento médio por vários examinadores; no trabalho cooperativo para a construção dos instrumentos de medida e, particularmente, no desenvolvimento da compreensão de que somente através do uso de padrões relativos de julgamento, com base no comportamento do grupo de examinados, que constitui o sistema referencial, é possível um julgamento adequado do desempenho escolar.

2.2 – O trabalho de elaboração dos instrumentos de medida é quase sempre realizado às pressas e sob pressão.

Há, implicitamente, um outro problema, que foge ao nosso campo de indagação, mas que não pode ser evitado, porque influi no estado atual das medidas educacionais. O construtor de instrumentos de medidas, geralmente professor militante, vê-se obrigado a exercer múltiplas atividades desgastantes, física e emocionalmente, e, assim, desconhece o lazer criativo, que lhe permitiria considerar a problemática do processo de orientar e avaliar a aprendizagem. Desse modo, os instrumentos de medida são elaborados às pressas, em véspera de exame e, às vezes, na própria hora do exame, sendo, portanto, instrumentos defeituosos e de má qualidade.

2.3. – Os instrumentos ora empregados são mal planejados e quase nunca possuem validade de conteúdo, por não considerarem uma amostra representativa de conhecimentos e capacidades relevantes.

Um instrumento de medida, qualquer que seja o seu aspecto formal, precisa ser adequadamente planejado. Um conjunto de 50 itens nem sempre é um teste, assim como uma dezena de perguntas de resposta livre nem sempre constitui uma prova de dissertação. A ausência de planejamento reflete-se em interrogação sensatamente apresentadas pelos examinados: — o que mede esse teste? qual o objetivo desse exame? o que pretende o professor com essas perguntas? Nem sempre é possível responder, com sinceridade e sem subterfúgios, a essas perguntas, que são justas e legítimas.

2.4 – Um grande número de instrumentos de medida enfatiza o trivial, o detalhe irrelevante, sem considerar capacidades educacionalmente importantes.

O importante, para alguns construtores de instrumentos de mensuração, é, por exemplo, a data do descobrimento da América, por Colombo, e não as características do impacto que esse descobrimento teve sobre a evolução da história do mundo moderno. Saber o nome de quem escreveu a obra *De Revolutionibus Orbium Coelestium* é mais relevante, felizmente

para um número reduzido de indivíduos, do que compreender o significado da revolução copernicana da ciência moderna. O conhecimento de elementos factuais e especiosos, entretanto, não é exclusivo de um único tipo de prova, parece ser uma tendência geral no atual estado dos instrumentos de medida.

2.5 – A estruturação formal inadequada de muitos instrumentos de medida não permite verificar capacidades complexas, como a de análise.

A deficiência formal é muitas vezes considerada por alguns críticos como sendo apanágio exclusivo dos itens objetivos. O argumento, ainda que sem apoio empírico, serve, frequentemente, para justificar e defender os itens de dissertação. Nada mais errôneo, pois, desde que bem elaborados, ambos os tipos de questão podem medir capacidades complexas. Há, na verdade, um problema técnico na estruturação de itens para a mensuração de capacidade complexas, que, infelizmente, nem todos os construtores de instrumentos conseguem solucionar satisfatoriamente.

A análise crítica de textos de dissertação mostra que as chamadas evidências da capacidade de analisar nada mais são do que exercícios de paráfrases de manuais e livros didáticos, sem nenhuma originalidade e profundidade. Entretanto, é um tipo de comportamento que pode ser verificado através de dissertações, desde que esse comportamento tenha sido desenvolvido durante o processo de aprendizagem. Ainda que bizarra, a situação realmente existe; muitas vezes, procura-se verificar comportamentos para os quais o examinando não recebeu treinamento prévio.

O problema é mais ou menos semelhante com relação aos testes objetivos. A falta de qualificação técnica de um construtor de itens não lhe permitirá elaborar uma unidade de informação que exija demonstração da capacidade de análise. Entretanto, a sua verificação é possível; assim, no caso de interpretação de textos literários ou científicos (Diederich, 1955) pode-se exigir que o examinando demonstre esse comportamento através da capacidade de:

- a – *identificar* objetivos e atitudes do autor do texto;
- b – *estabelecer* a idéia principal do texto;
- c – *mencionar* argumentos que apoiam a idéia principal;
- d – *assinalar* os pressupostos em que se baseiam as idéias do texto;
- e – *indicar* diferentes figuras de retórica;
- f – *criticar* a organização do texto;
- g – *julgar* a importância do texto;
- h – *avaliar* o mérito (literário ou científico) do texto.

2.6. – Os instrumentos de medida nem sempre são construídos com a observância dos princípios que devem orientar a sua construção.

Os defeitos de construção costumam ser mais aparente nos testes objetivos do que nos itens de dissertação. Entretanto, ao contrário da crença geral, um bom item de dissertação é coisa rara, pois é mais difícil de ser construído do que um item objetivo, sobretudo quando se pretende um instrumento de alta qualidade (Stanley, 1958).

Os itens objetivos, construídos sem a observância de normas técnicas, costumam apresentar os seguintes problemas:

- a – *desequilíbrio* na ênfase relativa das dimensões comportamento-conteúdo, o que reflete ausência de planejamento;
- b – *número reduzido* de questões, o que demonstra despreocupação com os problemas de validade e fidedignidade;
- c – *distribuição defeituosa* do índice estimado de dificuldade, geralmente obedecendo a padrões extremos de facilidade ou de dificuldade;

- d — irrelevância dos conhecimentos substantivos, inobservância da tecnologia do item e erros grosseiros de edição;
- e — favorecimento a padrões regulares de respostas, em virtude da falta de uma distribuição equilibrada das alternativas corretas;
- f — inconsistências gramaticais, falta de homogeneidade e plausibilidade das alternativas distratoras, o que favorece o acerto casual.

Os itens de dissertação também apresentam defeitos, ainda que menos visíveis para os não-especialistas, mas que nem por isso deixam de ser graves. Os vícios de construção mais freqüentes são os seguintes:

- I — proposição imprecisa dos temas ou das perguntas, caracterizada pela ausência de determinantes explicativos dos comportamentos exigidos, o que demonstra falta de planejamento;
- II — emprego indiscriminado da dissertação para verificar comportamentos que poderiam ser positivados por outros meios, o que traduz desconhecimento das várias técnicas de mensuração;
- III — irrelevância dos temas apresentados, que geralmente verificam comportamentos cognitivos simples;
- IV — número reduzido de questões, o que não permite um exame compreensivo e nem garante a validade do instrumento;
- V — ausência de critérios pré-estabelecidos e de padrões fixos de correção, que assegurem a fidedignidade dos escores;
- VI — influência de efeitos de halo e conseqüente contaminação dos escores, o que concorre para acentuar a subjetividade dos julgamentos e diminuir a precisão dos resultados.

2.7. — A maioria dos instrumentos de medida, ora usados no ambiente escolar, visam à avaliação somativa e não influem na orientação do processo de aprendizagem e nem informam com precisão sobre os supostos ganhos educacionais.

O emprego de instrumentos de mensuração educacional limita-se, freqüentemente, ao cumprimento de determinações administrativas, com a finalidade de “dar uma nota” e, após determinado período, aprovar ou reprovar, sem maiores preocupações docimológicas.

Os instrumentos nunca, ou quase nunca, são usados, por exemplo, no início de um curso ou de uma unidade, para fins de determinar o que o aluno sabe e, sobretudo, o que *não* sabe ou o comportamento que *não* possui, aspectos esses que permitiriam ao professor estabelecer um plano de trabalho para orientação da aprendizagem. Não há pré-testes e os exames finais não têm o caráter de pós-testes, evidentemente; desse modo, fica o professor impossibilitado de verificar se houve ganhos reais através do processo intencional da educação. Se os instrumentos de medida fossem adequadamente aplicados, poder-se-ia utilizar técnica apropriada (chi-quadrado) e, no caso, verificar a significância da diferença dos desempenhos e comprovar se a diferença resultou de efetiva modificação do desempenho escolar.

2.8. — Há um desconhecimento geral da influência exercida pelo tamanho do erro de amostragem nos escores de um teste.

Um teste objetivo ou uma prova de dissertação é uma amostra de conhecimentos e/ou comportamentos possíveis. Assim, qualquer que seja a forma do instrumento, a sua estruturação se faz com base em uma amostra selecionada segundo critérios fixados pelo examinador. O uso de amostras decorre de limitações óbvias, porquanto, no decorrer de um único exame, é impossível verificar o domínio de todos os conhecimentos e/ou a posse de todos os comportamentos possíveis.

Qualquer que seja o critério de seleção da amostra, inclusive no caso de uma amostra aleatória, comete-se um erro de amostragem. A magnitude desse erro está associada à não representatividade da amostra. Exames não compreensivos, baseados num número reduzido de

questões ou de itens, geralmente possuem um erro de amostragem grande. Um bom instrumento de medida procura atenuar a influência desse erro sobre o desempenho do estudante.

Os únicos erros, entretanto, que parecem preocupar são os que resultam do ajustamento incorreto da chave de correção ou do somatório de escores ou de notas. Há, assim, uma preocupação maior com erros decorrentes de distrações, que na verdade são enganos e podem ser evitados. Se não houve enganos e as somas estão corretas, os resultados são considerados precisos, em que pese a influência do erro de amostragem.

2.9. — A eficiência da maioria dos instrumentos de medida aplicados em nosso meio educacional não é verificada pela análise estatística.

A média e a variabilidade do grupo (desvio padrão) não são determinadas; o grau de dificuldade e o poder discriminativo (validade) dos itens ou questões não são estabelecidos; outros elementos necessários para a análise do instrumento, como o coeficiente de fidedignidade e o erro padrão de medida, não são calculados; entretanto, apesar de todas essas deficiências técnicas, os instrumentos são aplicados e decisões sobre o futuro dos estudantes são tomadas, enquanto que a maioria dos críticos se preocupa com aspectos formais e ignora outras implicações mais profundas que resultam da carência tecnológica de alguns instrumentos de medida.

3.0. — DIFERENTES MODELOS PARA A MENSURAÇÃO EDUCACIONAL.

Afirma-se algumas vezes, mas sem muita convicção, que estaria havendo, no momento, uma proliferação de testes objetivos, sobretudo nas escolas de 1^o e 2^o 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 de critério, testes de predição e testes combinados* (critério e predição).

Os testes de critério poderiam informar até que ponto os objetivos de uma unidade (ou curso) foram realmente alcançados e, conseqüentemente, possibilitariam 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; 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 de predição — a fim 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 predição — se completam por suas informações; contudo, um terceiro modelo poderia ser construído para obter os mesmos elementos que os outros proporcionam isoladamente. Ainda que mais complexo, o teste combinado é, na verdade, um teste de predição 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 predição 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 bens sucedidos têm a segunda parte do teste (predição) 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 5” —, 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 um novo exame — “a segunda época” —, geralmente tão duvidoso quanto os exames anteriores, e 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 predição, 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.

4.0. — QUALIFICAÇÕES TÉCNICAS DO CONSTRUTOR DE INSTRUMENTOS DE MEDIDA.

A verdadeira questão, no atual contexto educacional e no referente à avaliação do rendimento escolar, centraliza-se no fato de que muitos construtores de instrumentos de medida educacional não possuem a necessária formação técnica para o exercício de uma atividade específica que exige determinadas qualificações. Utilizam-se, às vezes, de uma tecnologia sofisticada, mas desconhecem os seus fundamentos teóricos.

As qualificações necessárias para o domínio da construção de instrumentos de medida educacional podem ser desenvolvidas através do:

- a — *conhecimento* das vantagens e das limitações dos atuais instrumentos de medida;
- b — *conhecimento* de critérios para o julgamento da qualidade dos instrumentos e dos meios de obter evidências relacionadas com esses critérios;
- c — *conhecimento* de como planejar um instrumento e elaborar diferentes tipos de itens ou questão;
- d — *conhecimento* de como aplicar eficientemente os instrumentos de medida;
- e — *conhecimento* de como interpretar corretamente os escores e outros elementos quantitativos.

5.0. — PROGRAMAS PARA O DESENVOLVIMENTO DE COMPETÊNCIAS TÉCNICAS.

O desenvolvimento de competências na área de tecnologia dos instrumentos de medida

permitirá garantir a validade do processo de avaliação educacional. Faz-se necessário evoluir da atual fase artesanal e ingressar na fase técnica, em que princípios científicos empiricamente estabelecidos substituem o espírito amadorista.

Um programa para esse fim poderia ser estruturado através:

- a – da intensificação dos currículos, na área de medidas educacionais, para a formação de professores, nas Faculdades de Educação;
- b – da criação e implementação de serviços de avaliação nas escolas de diferentes níveis, a fim de:
 - 1 – orientar professores na construção de instrumentos de avaliação;
 - 2 – definir objetivos educacionais relevantes e prioritários;
 - 3 – organizar programas de avaliação formativa e somativa;
 - 4 – determinar a eficiência do ensino e diagnosticar pontos críticos;
 - 5 – controlar a qualidade dos instrumentos construídos;
 - 6 – interpretar os resultados da aplicação dos instrumentos;
 - 7 – informar e orientar os estudantes sobre o seu desempenho escolar;
 - 8 – oferecer suporte administrativo para a elaboração e a aplicação de instrumentos de medida;
- c – da organização de programas especiais de curta duração, nas instituições educacionais, para a discussão de problemas, através de seminários e trabalhos práticos.

6.0. – RESUMO

1 – Os instrumentos de medida educacional, independentemente do seu aspecto formal, quando bem planejados e construídos, estimulam e orientam a aprendizagem do estudante.

2 – Qualquer que seja o tipo de instrumento, é necessário o domínio da tecnologia da sua construção, a fim de que sejam meios válidos de mensuração e fidedignos os resultados da sua aplicação.

3 – O sistema de mensuração ora em prática apresenta problemas que revelam a inobservância dos fundamentos teóricos e dos princípios tecnológicos que orientam a elaboração de instrumentos usados num programa de medidas.

4 – Apesar da existência de diferentes modelos para mensuração educacional, os mesmos não estão sendo utilizados em nenhum dos níveis de escolaridade.

5 – Através de um treinamento especializado, é possível desenvolver capacitações técnicas a fim de permitir a introdução e a implementação de programas válidos de medidas educacionais.

REFERÊNCIAS BIBLIOGRÁFICAS

- CRONBACH, L. J. (1970) – *Essentials of psychological testing*. 3rd edition. Nova Iorque; Harper and Row, Inc.
- DIEDERICH, P. B. (1955) – Making and using tests. *English Journal*. National Council of Teachers of English.
- EBEL, R. L. (1961) – Improving the competence of teachers in educational measurement. *The Clearing House*, 36, 2, October.
- EBEL, R. L. (1972) – *Essentials of educational measurement*. Englewood Cliffs, N J: Prentice-Hall, Inc.
- ENGELHART, M. D. (1964) – What to look for in a review of an achievement test. *Personnel and Guidance Journal*, 42, pages 616-19.
- KATZ, M. (1961) – Selecting an achievement test – principles and procedures. *Evaluation and Advisory Series, nº 3*. Educational Testing Service. Princeton, J. J.
- LAIDLAW, W. J. (1965) – Teacher-made test: models to serve specific needs. *The Clearing House*, February.
- STANLEY, J. C. (1958) – ABC's of test construction. *National Educators Association Journal*, April.

