

Testes centrados em critério: a sua incidência em educação

L. S. ALMEIDA / I. F. S. RIBEIRO / L. M. CORREIA *

Os testes centrados em critério não são utilizados com frequência entre os psicólogos portugueses (Almeida, Ribeiro & Cruz, 1989). Se a expressão «*criterion-referenced tests*» data já dos anos 60 (Glaser, 1963) e se existe já um grande número de instrumentos que se inserem nesta perspectiva noutros países, o facto é que os «testes centrados em critério»¹ não têm tido entre nós grande desenvolvimento. Constituído-se estes testes como complemento à avaliação formal clássica seria interessante a sua maior divulgação junto dos psicólogos portugueses². Registe-se, aliás, que também não possuímos em Portugal um número considerável, ou mesmo suficiente, de testes formais aferidos para os diferentes grupos etários e construtos psicológicos.

Neste artigo, procura apresentar-se genericamente estes novos testes e nova formulação da avaliação psicológica e, desta forma, contribuir para a sua maior difusão (informação) junto dos psicólogos portugueses.

* Professor Associado do Instituto de Educação da Universidade do Minho, Assistente do Instituto da Educação da Universidade do Minho, Escola Superior de Educação de Lisboa.

1 A designação destes testes não tem sido uniforme. Na parte final deste artigo referimos e exemplificamos esta dificuldade. Por outro lado, as traduções portuguesas oscilam mais frequentemente entre «testes centrados em critério», «testes com referência no critério» ou «testes referenciados em critério».

2 Alguns testes elaborados nesta perspectiva estão a ser usados embora geralmente confinados a alguns programas de inovação educativa ou alvos de pesquisa. A sua divulgação sistemática junto dos profissionais não é de facto frequente.

1. PERSPECTIVA HISTÓRICA: CARACTERÍSTICAS E FINALIDADES

Os testes centrados em critério têm a sua história. Inserem-se numa perspectiva de avaliação psicopedagógica não coincidente com a tradição psicométrica. Enquadram-se, sobretudo, na perspectiva comportamental ou behaviorista de avaliação e de intervenção psicopedagógica, onde cada comportamento é decomponível numa sequência de subcomportamentos isoláveis, susceptíveis de registo, com frequência directamente observáveis e onde o contexto e as leis da aprendizagem detêm um papel importante em tudo o que tem a ver com a sua ocorrência, manutenção ou extinção. Uma comparação com a abordagem psicométrica da avaliação parece-nos, pois, útil para uma melhor clarificação de conceitos.

Na abordagem tradicional da avaliação psicológica – perspectiva psicométrica ou diferencial – o comportamento era essencialmente definido, ou tido como condicionado, pelas variáveis internas aos indivíduos, caracterizadas por uma certa estabilidade temporal e situacional assumida *a priori* (Almeida, 1986). As diferenças encontradas no comportamento dos indivíduos eram então remetidas para as diferenças individuais nos próprios construtos. Por exemplo, o nível de sucesso escolar ou profissional dos sujeitos era explicado, pelo menos em parte, pelas suas capacidades intelectuais. Os testes psicológicos surgiram como meio, por exce-

lência, para a avaliação quer das características intra-individuais quer das diferenças interindividuais. A aplicação dos testes junto de amostras representativas de indivíduos permitiam a obtenção de *normas de grupo* (média, desvio-padrão, notas percentílicas) para efeitos de interpretação dos resultados em aplicações posteriores dos mesmos testes. Daí a expressão «testes centrados em normas» (*norm-referenced tests*).

O aparecimento dos testes centrados em critério deu-se no domínio da educação, domínio onde aliás permanecem ainda hoje mais desenvolvidos, como alternativa aos testes formais de inteligência e de realização. Estes não se adequavam nem à avaliação da eficácia dos programas de promoção educativa então em voga nem à avaliação de crianças com necessidades educativas especiais. A sua inadequação era resultado de vários factores. Primeiro tratava-se de medidas ou variáveis não directamente ligadas aos conteúdos dos próprios programas e, como tal, pouco passíveis de fornecer pistas para uma melhor compreensão das necessidades e melhor programação, individualização ou rentabilização do próprio ensino. Segundo, ao assumirem a «normalidade» da distribuição dos resultados em tais características psicológicas e a necessidade dos resultados nos testes corresponderem a essa suposição, os itens estavam sobretudo orientados para a diferenciação dos sujeitos entre si. Os cuidados metodológicos (poder discriminativo, índice de dificuldade dos itens, sensibilidade dos resultados) na construção dos testes formais obedeciam a tal preocupação. O objectivo da avaliação tradicional era o de determinar a posição relativa de um indivíduo em relação aos seus pares, tendo em vista uma decisão face a várias alternativas que se lhe colocavam (tipo de ensino, orientação vocacional, selecção profissional). Esta distribuição normal dos resultados nem sempre é desejável. Por exemplo, após um programa de treino ou no ensino em geral espera-se que a maioria ou a quase totalidade dos indivíduos obtenham sucesso.

Os testes formais não se adequavam aos objectivos agora prosseguidos. As tentativas nos anos 60 de utilização dos testes formais ou testes centrados em normas para a avaliação da eficácia dos programas educativos geraram, então, várias reacções negativas por parte de educadores e psicólogos mais directamente envolvidos em tais acções. Os *testes referenciados em critério* surgem neste contexto

como a alternativa: era necessário encontrar instrumentos sensíveis aos efeitos dos programas educativos que nos anos 60 e 70 se multiplicavam, dando origem a modelos como o ensino individualizado, a aprendizagem para a mestria, a sequência hierárquica de ensino, o ensino assistido por computador, o ensino programado e a outros modelos associados aos conceitos de avaliação somativa e formativa. Nos testes formais o problema não estava apenas na sua baixa sensibilidade às aquisições efectuadas (aliás pouco identificados com os construtos avaliados em tais testes), como também na tentativa de distribuição dos sujeitos à luz dos princípios estatísticos (leis da curva normal). Diríamos que, independentemente dos programas, deveria haver uma distribuição normalizada e equitativa de sujeitos abaixo e acima da média. Este último aspecto encontra-se, à partida, em oposição clara com a defesa de uma educação assente na igualdade de oportunidades para todos.

A nova filosofia do *testing* não está centrada na diferenciação dos sujeitos mas nos próprios objectivos de intervenção. Pretende, apenas, verificar-se em que medida os sujeitos atingem ou não determinado nível de realização em determinado conteúdo ou critério, ou seja, aspira-se à descrição da sua realização em termos de comportamentos específicos manifestados. Se aceitarmos, por exemplo, que a prática educativa justifica uma programação dos conteúdos (sequência) e uma individualização do processo ensino-aprendizagem, reconheceremos que este tipo de testes serve mais facilmente esse duplo objectivo. Por um lado, delimita-se um domínio e presta-se uma melhor atenção à sequência e à justificação dos respectivos conteúdos. Por outro, e no quadro de uma sequência instrucional, permite a monitorização da aprendizagem dos indivíduos e a fixação de níveis de mestria para a sua prossecução. A partir da análise das tarefas e das condições, a aprendizagem pode ser pensada como uma progressão hierarquizada e feita por unidades ou passos sucessivos. Podemos afirmar que os testes referenciados em critério se enquadram perfeitamente no que vem sendo a pedagogia por objectivos (identificar os objectivos, operacionalizar o que há a ensinar, adequar as estratégias às aquisições e ao ritmo de aprendizagem dos alunos, avaliar o que está ou o que falta aprender, reavaliar e retomar este processo), devendo, de acordo com Glaser e Nitko (1971), ser construídos deliberada-

mente para produzir resultados que sejam directamente interpretáveis em termos de padrões de realização específicos.

Com este tipo de fundamentação teórico-prática, não faz sentido nestes testes o recurso às realizações individuais para a partir daí se formularem as normas de interpretação dos níveis atingidos. Fará sentido, sim, definir previamente o nível a atingir e orientar a programação educativa para que todos os sujeitos o satisfaçam. Ao contrário dos testes centrados em normas, onde os critérios são relativos e *a posteriori*, nos testes centrados em critério tais níveis são definidos com objectividade e prévios à realização dos sujeitos. Eles devem permitir, aliás, orientar a intervenção de forma mais objectiva (por exemplo através da reformulação das estratégias ou dos próprios objectivos).

Mais que informar se um determinado aluno se situa no grupo dos que provavelmente têm 25% de probabilidades de sucesso (probabilidades calculadas com base nas tabelas de expectativas nos testes formais), os novos testes permitem um conhecimento *aqui e agora* daquilo que o sujeito consegue fazer e da forma como isso pode ser utilizado para promover as suas aprendizagens. Trata-se de testes sobretudo prescritivos, enquanto os testes tradicionais valorizavam acima de tudo a sua capacidade preditiva ou prognóstica. Em termos práticos os testes centrados em critério, mais na linha dos «testes edumétricos» (Craver, 1974), dão uma informação importante quando está em causa o desenvolvimento de objectivos educativos específicos para cada criança ou grupos de crianças, a avaliação de eficácia das estratégias educativas implementadas para se atingir os objectivos educativos ou a própria avaliação dos objectivos inicialmente formulados.

Possibilitam, por outras palavras, o fácil estabelecimento de uma ponte entre o diagnóstico e a remediação das dificuldades a cada momento experienciadas. No caso da criança com necessidades educativas especiais, para a qual uma parte substancial do ensino é individualizado, os testes centrados em critério, pela sua especificidade, dão-nos uma informação *exacta*, directamente aplicável na elaboração de programas de intervenção.

Em resumo, podemos situar as características principais deste modelo de *testing* a três níveis: apreciação da realização individual sem referência à realização dos outros indivíduos, tarefas referentes

a um domínio bem delimitado de realização e classificação dos sujeitos apenas em termos de «mestria» ou não «mestria» no critério avaliado (Howell, Kaplan & O'Connell, 1979 Black, 1985). Tais características granjearam para estes testes uma grande popularidade no momento presente, dada a sua incidência educacional (fácil adequação destes testes às situações individuais de aprendizagem, facilidade na sua construção, fácil inserção nos objectivos dos programas ou adequação no tempo aos programas em acção), permitindo ainda situar a criança de acordo com as suas áreas fortes e fracas em oposição à comparação com um grupo ou classe. Assim, dado que a ênfase está nas aptidões e não na comparação, determinar o *que se* deve ensinar e *como* avaliar esse ensino torna-se mais simples.

O Quadro 1, adaptado de Ahmann e Glock (1975), dá-nos uma ideia das diferenças e semelhanças existentes entre os testes centrados em normas e os testes centrados em critérios, incorporando muitas das ideias expressas neste artigo.

2. CARACTERÍSTICAS METROLÓGICAS DOS RESULTADOS

A especificidade destes testes levanta sérias dificuldades à utilização dos parâmetros estatísticos na tradição psicométrica para apreciação das características metrológicas dos resultados. Aqui a apreciação terá que incidir mais na análise dos itens e na relevância dos resultados para os objectivos educacionais em causa do que no cálculo de coeficientes estatísticos. Estes, dada a sua natureza correlacional, estão de alguma forma afectados pela não valorização da diferenciação dos sujeitos neste tipo de testes e pelo baixo número de itens em média. Repetimos, a favor dessa não diferenciação esperada, que a educação nos nossos dias toma como objectivo não a seriação e a selecção dos alunos, mas que todos atinjam os parâmetros fixados.

Estas considerações não impedem que haja alguns cuidados com a construção destes instrumentos (formato dos itens, significância dos itens, sintaxe e semântica das questões) e com a sua utilização (condições externas da avaliação, variáveis dos sujeitos e do avaliador). Algumas exigências devem ser respeitadas no momento da construção. Em primeiro lugar é indispensável a definição clara do domínio e das tarefas da aprendizagem a avaliar.

QUADRO 1

Comparação de Testes Centrados em Normas com Testes Centrados em Critérios

Características do teste	Testes centrados em normas	Testes centrados em critérios
Planeamento do teste	O teste é geralmente planeado com base em descrições gerais de tópicos globais e capacidades intelectuais.	O teste é geralmente planeado tendo em conta objectivos educacionais em termos de comportamento, podendo de cada um deles derivar um ou vários itens de avaliação.
Preparação dos itens	Os itens são elaborados com o fim de maximizar a discriminação entre os sujeitos.	Os itens são elaborados com o fim de avaliarem a proficiência do sujeito em determinada tarefa.
Tipos de itens	Todos os tipos	Todos os tipos
Nível de dificuldade dos itens	Os itens são de dificuldade média, ou seja, cerca de 30 a 70 por cento dos sujeitos responde correctamente.	A dificuldade dos itens varia, embora geralmente sejam do tipo de «mestria», ou seja, a maioria dos sujeitos responde correctamente depois de ser treinada.
Critério de mestria	Geralmente não se especifica qualquer critério de mestria.	Os objectivos dos itens designam um critério para a mestria, ou muitas das vezes ele pode ser inferido.
Interpretação dos resultados	Trata-se de um resultado (ou vários resultados) para a determinação da posição relativa do sujeito em relação a um grupo.	Determina-se o sucesso ou insucesso do sujeito num item ou pequeno grupo de itens similares, descrevendo-se de seguida, a sua realização com referência a determinados objectivos.
Finalidades do teste	<ol style="list-style-type: none"> 1. Determinar a posição relativa do sujeito num grupo normativo. 2. Auxiliar na selecção de indivíduos para programas educacionais, de emprego, etc. 3. Dar uma ideia global do aproveitamento do aluno numa determinada área de aprendizagem. 	<ol style="list-style-type: none"> 1. Determinar a posição do sujeito em relação a padrões de realização estabelecidos, precisando-se, assim, o que ele sabe e é capaz de fazer. 2. Diagnosticar problemas específicos de aprendizagem. 3. Avaliar o sucesso de um programa de ensino.

Esta exigência tem a ver não só com a delimitação dos conteúdos ou unidades programáticas, como também com a explicitação prévia do tipo de resultados esperados (aquisição, compreensão, aplicação, generalização... e tudo isto em relação a conceitos, a princípios ou a procedimentos, por exemplo). Uma segunda exigência tem a ver com a necessidade de uma definição dos objectivos educacionais em termos de comportamentos ou tarefas precisas a realizar. Tais itens a integrar o teste tornam-se, então, importantes quer na programação das actividades educativas quer na avaliação do seu impacto.

Por último, é necessário fixar previamente os níveis de realização tidos como desejáveis. Este aspecto difere claramente dos procedimentos utilizados nos testes centrados em normas onde os níveis de interpretação dos resultados são calculados *a posteriori* tomando as realizações dos indivíduos. Aqui, os níveis são fixados previamente numa planificação educativa e enquadrados nos objectivos educacionais mais latos. Tais níveis ou padrões poderão, aliás, não ser uniformes para as diversas unidades programáticas ou para os diferentes tipos de aprendizagem a promover. A elaboração prévia de tabelas taxonómicas e de especificação dos itens permite uma atenção simultânea a estes vários aspectos.

Ao nível da fidelidade alguns autores sugerem o uso da metodologia do teste-reteste para apreciação da confiança dos resultados obtidos. Assim, aplicado o mesmo teste ou forma paralela aos sujeitos em dois momentos diferentes, geralmente com um intervalo de duas semanas, calcula-se a percentagem de acordo entre as classificações individuais de mestria e de não-mestria nas duas aplicações. Subentende-se que entre as duas aplicações do teste não há intervenção deliberada no domínio avaliado. Por outras palavras, trata-se tão somente de um estudo para avaliar as características metro-lógicas dos resultados na prova em questão. Assim, na metodologia do teste-reteste, pretende avaliar-se a constância quer do construto a medir e do teste em si como instrumento de diagnóstico quer das características do sujeito.

Ao nível da análise da validade, os itens a integrar este novo tipo de testes devem representar convenientemente o domínio em avaliação ou o critério (validade de conteúdo). Só desta forma podemos pensar que o teste em questão serve para avaliar as competências dos sujeitos nesse critério.

Ao mesmo tempo, os itens devem ser capazes de prever o sucesso subsequente (validade externa) dos alunos. Por exemplo, estando estes testes ligados a uma planificação do ensino por módulos, é esperado que a entrada num determinado módulo apenas ocorra quando o aluno atinge a mestria no módulo anterior. Assim sendo, importa que não apenas os itens representem o módulo em questão como se constituam como bons indicadores do sucesso dos alunos nas aprendizagens subsequentes. Aspecto fundamental para um e outro tipo de validade é a salvaguarda prévia da pertinência e clara definição dos conteúdos e dos próprios comportamentos a integrar no teste. Nesta altura, situamo-nos já naquilo que os autores definem por validade de construto, embora com bastantes afinidades com a validade de conteúdo. Um teste bem construído pode dar-nos um manancial de informação acerca do reportório comportamental de determinada criança. Possibilita-nos, assim, descobrir o que ela é capaz de fazer, isto é, permite-nos verificar se ela beneficiou ou não de alguns dos resultados imediatos da situação de aprendizagem.

3. CONSIDERAÇÕES FINAIS: ALGUMAS DIFICULDADES

A concluir, poder-se-ão apontar algumas das dificuldades que este tipo de testes apresenta. Em primeiro lugar existe alguma confusão quanto à sua designação. Por exemplo, vários termos aparecem frequentemente usados como sinónimos de «testes centrados em critério». Estão neste caso as expressões *mastery tests* (Mayo, 1970), onde a ênfase aparece colocada no nível de realização, *domain-referenced tests*, expressão bastante próxima e onde se salienta o conteúdo ou âmbito, e *objective-referenced tests* (Ebel, 1979) onde se enfatiza o carácter comportamental dos itens a formular. A expressão *criterion-referenced tests* parece ser assumida pelos autores como a mais correcta ou, pelo menos, objecto de maior preferência.

Uma segunda dificuldade advém do duplo significado que o termo «critério» parece, por vezes, assumir na literatura sobre o assunto. Alguns autores tomaram-no inicialmente como sinónimo de nível ou padrão de realização. Os autores tendem hoje a clarificar que «critério» deve ser identificado com o

domínio, conteúdo ou âmbito das situações a ensinar e a avaliar. Em relação ao nível de realização, os autores preferem usar a expressão «mestria».

Quanto ao nível de realização a fixar como «mestria» num determinado teste também não existem parâmetros seguros de orientação. É possível que os mesmos se tenham de adequar ao próprio objectivo da avaliação (objectivos diferentes de uma avaliação de processo para uma avaliação final ou de certificação). Block (1971) sugere que em termos educacionais tal nível deve situar-se entre 80% e 85% de respostas correctas. Aspectos motivacionais dos sujeitos justificariam tal proposta, apesar de sabermos que a mesma não pode também ser tomada como independente do formato dos itens no teste (por exemplo nas questões de escolha múltipla o resultado dos sujeitos aparece afectado pela margem de acaso das suas respostas, aumentando-se a probabilidade de respostas ocasionais correctas à medida que se diminui o número de alternativas para opção dos sujeitos (Almeida, 1989).

Registe-se, ainda, que o sentido de «mestria» na avaliação do comportamento tem sido erradamente tomado como sinónimo de percentagem de respostas correctas. Trata-se de um reducionismo, pois que também pode ser analisada em termos de velocidade (ser capaz de realizar a tarefa em determinada porção de tempo) e em termos de precisão da realização (grau de perfeição exigida ou conseguida).

Por último, colocar-se-ia a questão da total suficiência destes testes para as diversas situações da prática educacional, nomeadamente para a intervenção psicológica. Aqui a resposta mais frequente aponta para a complementaridade da informação obtida através dos testes centrados em critério e dos testes centrados em normas. A promoção da aprendizagem, a orientação educativa e a orientação vocacional, por exemplo, poderão requerer em determinados momentos os dois tipos de informação. Em determinadas situações poderá ser necessário conhecer os níveis individuais de realização em determinadas competências cognitivas básicas, os seus níveis ou estádios de desenvolvimento, e a posição relativa dos sujeitos em relação aos seus pares. Por seu turno, as avaliações educacionais (inicial, formativa, sumativa) e a própria monitorização do ensino requerem os testes referenciados em critério (Gronlund, 1973, 42-5). Mais focalizados nos próprios objectivos educativos estes últimos mostram-se particularmente úteis no diag-

nóstico de necessidades, na avaliação do processo de ensino-aprendizagem, na certificação da aprendizagem, na avaliação da eficácia de programas de intervenção, na avaliação individual das competências em diferentes currículos, na avaliação dos objectivos fixados ou dos métodos e estratégias implementados.

REFERÊNCIAS

- AHMANN, J. S. e GLOCK, M. D. (1975) — *Evaluating pupil growth: Principles of tests and measurement*, Boston: Allyn and Bacon, Inc.
- ALMEIDA, L. S. (1986) — «Inteligência: Evolução no seu estudo», *Jornal de Psicologia*, 5 (3), 13-17.
- ALMEIDA, L. S. (1989) — «As respostas ao acaso nos testes de aptidão: estudo segundo o sexo e o ano de escolaridade», in J. F. Cruz, R. A. Gonçalves e P. P. Machado (Eds), *Psicologia e Educação: investigação e intervenção*.
- ALMEIDA, L. S.; RIBEIRO, I. S. e CRUZ, O. (1989) — «Uso dos testes no contexto da avaliação psicológica: resultados preliminares. Poster apresentado na Conferência Internacional «A Psicologia e os Psicólogos Hoje», Lisboa: APPORT.
- BLACK, H. D. (1985) — «Wither research on criterion-referenced assessment?», *Research Intelligence*, 19, 2-5.
- BLOCK, J. D. (1985) — *Mastery learning: Theory and practice*, New York: Holt, Rinehart & Winston.
- CARVER, R. P. (1974) — «Two dimensions of tests: Psychometric and edumetric», *American Psychologist*, 29, 512-518.
- EBEL, R. L. (1979) — *Essentials of educational measurement*, (3rd ed), Englewood Cliffs, N. J.: Prentice-Hall, Inc.
- GLASER, R. (1963) — «Instructional technology and the measuring of learning outcomes: Some questions», *American Psychologists*, 18, 519-521.
- GLASER, R. e NITKO, A. J. (1971) — «Measurement in learning and instruction», in *Educational measurement*, ed. R. L. Thorndike Washington, D. C.: American Council on Education, 635-670.
- GRONLUND, N. E. (1973) — *Preparing criterion-referenced test for classroom instruction*, New York: MacMillan.
- HOWELL, K. W., KAPLAN, J. S. e O'CONNELL, C. Y. (1979) — *Evaluating exceptional children — a task analysis approach*, Columbus, OH: Charles E. Merrill Publishing Company.
- MAYO, S. T. (1970) — «Mastery learning and mastery testing National Council on Measurement in Education (NCME)», *Measurement in Education*, 1,1-4.

RESUMO

Este artigo apresenta alguns aspectos que, nas duas últimas décadas, têm tornado os testes centrados em critério como uma alternativa aos testes formais, particularmente no domínio da educação. A par das principais características e finalidades destes testes, referem-se as dificuldades de uso da metodologia estatística tradicional na sua construção ou na apreciação da fidelidade e validade dos seus resultados. Por último, aponta algumas dificuldades específicas destes instrumentos e a complementaridade que na intervenção psicopedagógica os testes centrados em critério e os testes centrados em normas podem representar.

ABSTRACT

This article presents some aspects that in the last two decades have turned the criterion-referenced tests into an alternative to traditional tests, particularly in the educational field. Furthermore, it refers to the main characteristics and objectives of these tests, focuses the use of traditional statistical methodology in their development and analyses the reliability and validity of their scores. Finally, it presents some specific difficulties common complementary nature as tools used in a psychoeducational intervention.