·EDUCAÇÃO HOJE·

O CRITÉRIO DO SUCESSO

TÉCNICAS DE AVALIAÇÃO DA APRENDIZAGEM

Valter V. Lemos

Texto Editora

TÍTULO O CRITÉRIO DO SUCESSO — TÉCNICAS DE AVALIAÇÃO DA APRENDIZAGEM
AUTOR VALTER V. LEMOS
CAPA GABINETE DE PROJECTOS GRÁFICOS TEXTO — Cristina Malaquias
ARRANJO GRÁFICO GABINETE TÉCNICO TEXTO
EDITOR TEXTO EDITORA, LDA.
COMPOSIÇÃO EUROCOMP — GABINETE DE FOTOCOMPOSIÇÃO E ARTES GRÁFICAS, LDA.
MONTAGEM SECTOR GRÁFICO TEXTO
IMPRESSÃO E ACABAMENTOS EMPRESA INDUSTRIAL DE FOTOLITOGRAFIA, LDA.
© TEXTO, LDA. LISBOA, 1990 • 4.ª EDIÇÃO
Depósito legal n.º **39606/90**

© 1990, TEXTO EDITORA, LDA.

Reservados todos os direitos. É proibida a reprodução desta obra por qualquer meio (fotocópia, offset, fotografia, etc.) sem o consentimento escrito da Editora, abrangendo esta proibição o texto, a ilustração e o arranjo gráfico. A violação destas regras será passível de procedimento judicial, de acordo com o estipulado no Código do Direito de Autor e dos Direitos Conexos.

O verdadeiro mistério do mundo é o visível e não o invisível.
Oscar Wilde

As teorias passam. A rã fica.
Jean Rostand

Fiquei aliviado com esta súbita passagem do geral para o particular.
Sherlock Holmes

Pág.

CAPÍTULO IV — TRATAMENTO DE RESULTADOS DE TESTES E QUESTIONÁRIOS 47
Medidas de tendência central 49
Moda 49
Mediana 50
Média aritmética 51
Medidas de dispersão 52
Desvio-padrão 52
Bibliografia 54

CAPÍTULO V — A CORRECÇÃO E PONTUAÇÃO DOS TESTES E A CLASSIFICAÇÃO DOS ALUNOS 55
Correcção e pontuação de resultados 57
Cotação e escalas 57
Critérios de sucesso e pontuação 59
Correcção e pontuação de testes objectivos 59
Correcção e pontuação de testes não-objectivos 60
Correcção e pontuação de registos de observação 62
Classificação dos alunos 64
Utilização da escala 1 a 5 65
Utilização da escala 0 a 20 66
Bibliografia 67

CAPÍTULO VI — A UTILIZAÇÃO DOS DADOS DE AVALIAÇÃO DA APRENDIZAGEM — A AVALIAÇÃO DO ENSINO 69
Orientação e aconselhamento 71
Recuperação e enriquecimento 72
Modificação dos métodos de ensino 73
Avaliação de programas e investigação 75
Bibliografia 77

ÍNDICE

Pág.

CAPÍTULO I — AVALIAÇÃO: TEORIA E ACÇÃO 9
Introdução 11
As funções da avaliação 14
Avaliação criterial e avaliação normativa 15
Níveis de aprendizagem 18
Resumo 20
Bibliografia 21

CAPÍTULO II — A CONSTRUÇÃO E UTILIZAÇÃO DOS INSTRUMENTOS DE AVALIAÇÃO 23
Técnicas de Avaliação (Quadro I) 25
Os testes — Tipos e caracterização 26
Testes-itens não objectivos 28
Testes-itens objectivos 28
Resposta-curta 28
Verdadeiro/Falso 29
Combinação 30
Escolha múltipla 31
Registos de observação 32
Listas de verificação 32
Grelhas de análise 33
Questionários de opinião 34
Bibliografia 35

CAPÍTULO III — QUALIDADE DOS TESTES 37
Validade 39
A validade de conteúdo 40
Diferencialidade 41
Grau de dificuldade 41
Índice de discriminação 42
Utilização dos índices de dificuldade e discriminação 43
Fidelidade 44
Objectividade 45
Bibliografia 46

"Quem mais conhece melhor ama"

Da experiência de trabalho na formação de professores, quer na formação inicial, quer na formação em serviço, quer na especialização, verifiquei a menor importância que muitos professores dão ao *rigor*, na medição da aprendizagem dos seus alunos.

Para além desta ser a componente mais importante da avaliação do ensino e de todo o sistema de ensino-aprendizagem, acresce a necessidade de os professores classificarem numa escala os seus alunos e dessa classificação depender, além de outras coisas, o sucesso dos mesmos alunos. Isto implica forçosamente o maior rigor na avaliação da aprendizagem, por razões de ordem pedagógica, psicológica e social. *Será absolutamente incompreensível e inaceitável que um aluno possa ser condenado ao insucesso por razões exteriores à aprendizagem, tais como um teste inadequado ou um erro estatístico do professor.*

Sabe-se que as razões do insucesso escolar são múltiplas e se prendem com alguns factores não controláveis pelo professor. Mas, a verdade, é que também não se têm melhorado suficientemente alguns aspectos da acção escolar, que são perfeitamente controláveis e susceptíveis de aperfeiçoamento. No âmbito da avaliação tem proliferado a literatura sobre a discussão teórica dos conceitos de avaliação e das suas funções, rareando a respeitante a instruções e metodologias práticas de trabalho.

É evidente que a forma como se realiza a avaliação não é absolutamente independente das teorias em que se baseia a estruturação do ensino de cada professor. Parece-me, no entanto, que a necessidade do rigor é condição generalizada. Eu próprio, já usei diversos quadros teóricos na estruturação da prática escolar: a concepção beaviorista de aprendizagem hierarquizada, a pedagogia para a mestria, a estruturação conceptualista e a teoria cognitivista. No entanto, as diversas práticas, aliadas a estudos e investigações realizadas, mais mostraram a clara necessidade do professor avaliar com o maior rigor possível a aprendizagem dos seus alunos.

Todos sabemos que a aprendizagem escolar não diz respeito somente às actividades cognitivas, abrangendo outras áreas com a afectiva, a social, a motora, etc., mas, a realidade mostra que é àquela que os professores, nas nossas escolas, dão o papel determinante. E a regra de que ao *ênfase no ensino deve corresponder igual ênfase na avaliação* é pedagógica, psicológica e socialmente indiscutível. Claro que é necessário desenvolver métodos e instrumentos adequados à aprendizagem respeitante a esses campos (e neste trabalho abordam-se alguns), no entanto, a existência desses aspectos que não podemos ou não somos capazes de avaliar, não é razão para não avaliar bem o que podemos. Pretender que o que é avaliado, o seja da forma mais adequada e rigorosa possível, é um objectivo fundamental de qualquer sistema de ensino-aprendizagem.

Em 1979 fiz parte de uma equipa, que, informal, mas rigorosamente, analisou o enunciado de algumas provas nacionais de exame do curso complementar do ensino secundário, em disciplinas da área de ciências. Numa dessas provas, num total de 25 itens, 17 continham erros de forma ou de conteúdo. Penso que a descrição do facto dispensa comentários.

Liderava essa equipa, o Dr. José Barreiros Mateus. A ele, o mais profundo agradecimento, pela postura mais profissional que já conheci, e por todas as ideias e trabalhos que desenvolvemos juntos. Um agradecimento também a todos os que, até hoje, comigo trabalharam nesta difícil, mas aliciante profissão de professor, e que têm também, de algum modo, uma quota-parte neste pequeno trabalho, porque, se no final do século XX, alguém pode fazer alguma coisa, completamente sozinho, não é, de certeza, «fazer educação».

1

AVALIAÇÃO, TEORIA E ACÇÃO

Introdução

Desde há alguns anos instalou-se em Portugal no domínio da formação em serviço (estágios, prática, formação em exercício, etc.) dos professores dos ensinos preparatório e secundário, a norma de planificação de acordo com as ideias de B. S. Bloom e colaboradores. A utilização que delas foi feita redundou, no entanto, em quase completo desastre, pois que acabou por produzir uma reacção bastante negativa por parte de grande número de professores que após um ano ou dois de submissão (por vezes tirânica) a processos de planificação que desconheciam e que nunca conseguiram ver explicados, pelo menos no respeitante às suas vantagens de aplicação prática, acabaram por arrumá-los na prateleira do esquecimento, logo após a conclusão da sua formação.

As consequências tornam-se ainda mais graves, porque muitos desses professores acabam por criar um sentimento de rejeição não só às ideias e processos de B. S. Bloom, conforme estas lhes foram apresentadas, mas também a todo e qualquer processo de planeamento estruturado do sistema ensino-aprendizagem.

Curiosamente, e apesar desta insistência nas ideias de Bloom, no tipo de formação pedagógica que maior número de professores tem abrangido, a edição portuguesa de obras daquele autor quase não existe (divulgando-se alguma brasileira) e a produção portuguesa de trabalhos de tratamento e divulgação das suas teorias, tem representação diminuta.

Poderemos até dizer que as obras fundamentais de estruturação teórica, do autor, *"Human characteristics and school learning"* e *"Handbook on formative and summative evaluation of student learning"*, praticamente não circulam em Portugal.

Neste panorama, não é difícil descortinar as razões do fracasso da implantação real dos processos de organização, desenvolvimento curricular e avaliação propostos por Bloom e colaboradores.

No entanto, para além disso uma razão fundamental existe para tais acontecimentos e que não se cinge à teoria referida, mas a uma grave lacuna no processo educativo em Portugal. Trata-se do rigor absolutamente imprescindível no que respeita à avaliação da aprendizagem dos alunos e cujo desenvolvimento e aplicação é descurado há bastantes anos.

Na realidade, o problema não é especificamente português, rareando mesmo nos países desenvolvidos a produção de trabalhos exclusivamente dedicados a este tema. Se procurarmos com atenção, verificamos que do relativamente pequeno número de

publicações nesta área, um número ainda significativamente mais pequeno tem completo interesse e aplicabilidade pelos professores.

A situação deriva fundamentalmente da razão de grande parte das obras estarem direccionadas, ou pelo menos, claramente influenciadas por aspectos relacionados com a avaliação de carácter psicológico e não pedagógico. Ao que pensamos, tal dever-se-á a duas características fundamentais do campo psicológico, que facilitaram o desenvolvimento de tais metodologias: o carácter de «normalidade» e standartização da avaliação psicológica e o desenvolvimento desta área do conhecimento de forma científica e clínica.

Efectivamente a distribuição estatística «normal» de uma característica psicológica numa população, torna mais fácil a construção e validação de instrumentos de avaliação dessa característica. Isto permitiu desenvolver e aperfeiçoar uma tecnologia de avaliação com segurança estatística.

No caso do ensino e da aprendizagem a situação foi sempre mais difícil, dado não poder falar-se, do mesmo modo em «distribuição normal», pois que não se trata da *distribuição aleatória*, de uma característica numa população, mas sim de uma *distribuição, que se pretende não aleatória.*

Assim, a standardização de instrumentos, bem como os processos de validação (dado que o professor, geralmente, não tem condições de aplicação prévia para validação) tornam-se mais difíceis, o que foi permitindo a «colonização» metodológica que, no entanto, não tem resolvido, por completo, os problemas da avaliação pedagógica.

Com o desenvolvimento das ideias de «diferenciação» e de «individualização» no ensino e aprendizagem, mais o problema se agudizou, tendo, então, que se ter em conta já não a chamada avaliação em função da «norma», mas sim em função do «critério» previamente estabelecido para a aprendizagem desejada.

É esta a principal consequência do aparecimento da chamada «pedagogia para a mestria», tantas vezes confundida e designada pela redundância de «pedagogia por objectivos» (haverá pedagogia sem objectivos?), que aliás, também tem sido usada para designar outras teorias do ensino.

Há, na realidade, aspectos no sistema ensino-aprendizagem que podem referir-se à «norma» em termos de avaliação e outros que podem referir-se ao «critério». É importante ter os dois em consideração, identificar as situações e usar as metodologias mais adequadas em cada caso.

A influência da avaliação psicológica na avaliação pedagógica teve ainda outra consequência importante: a dificuldade de encontrar regras práticas que respondes-

sem às características de caso único (no tempo, relativamente ao mesmo aluno) dos instrumentos de avaliação de aprendizagem. É aqui, aliás que reside a diferença fundamental: *enquanto em psicologia os instrumentos de avaliação se destinam a medir características de existência geral em toda a população em qualquer momento temporal, em pedagogia destinam-se a medir capacidades de realização ou conhecimentos de existência individual ou de pequenos grupos, num momento determinado no tempo.*

Tudo isto fez com que o estudo teórico e prático sobre os instrumentos de avaliação da aprendizagem, tenha sido reduzido, durante muito tempo, na perspectiva essencialmente pedagógica.

Verificamos assim, hoje, a proliferação de literatura sobre avaliação, que analisando aspectos teóricos e conceptuais sobre fundamentos e tipologias, muito pouco tem dito sobre aspectos instrumentais e operacionais. Se a fundamentação é um pressuposto essencial de acção, a instrumentação é, ela própria, uma componente da própria acção.

Dizer que tal aspecto tem sido descuidado nas ciências de educação em Portugal, é evidentemente redundante, sabendo todos os professores as pouquíssimas indicações que são dadas sobre construção e utilização de testes, grelhas, questionários, etc., nos inúmeros cursos e seminários sobre avaliação que têm proliferado, e nos quais, em compensação, muito se diz sobre avaliação normativa e criterial, formativa e sumativa, etc., etc..

Evidentemente que é necessário esse tipo de enquadramento sempre que se trate de avaliação, e dele já se falou atrás e ainda se irá falar um pouco mais adiante. Mas o objectivo fundamental deste trabalho é, divulgar de forma simples e clara, regras e orientações para a elaboração de instrumentos de avaliação da aprendizagem, que possam ser usados no trabalho de todos os dias, quer como fonte orientadora, quer como fonte de recurso para certificação do trabalho já elaborado.

Não se pretende criar nada de especialmente novo, mas, somente, sistematizar de modo simples e operacional princípios e regras já conhecidas, mas muitas vezes não sedimentadas ou associadas. Trata-se, portanto, também de um instrumento de acção, que arriscando a crítica sobre a insuficiente análise e justificação teórica, se pensa mais necessário e útil aos professores, na presente forma.

As funções da avaliação

O sistema de ensino-aprendizagem é composto por três partes fundamentais que têm forma sequencial, mas que se influenciam também de forma retroactiva. Podemos representar isto num esquema simples (Fig. 1).

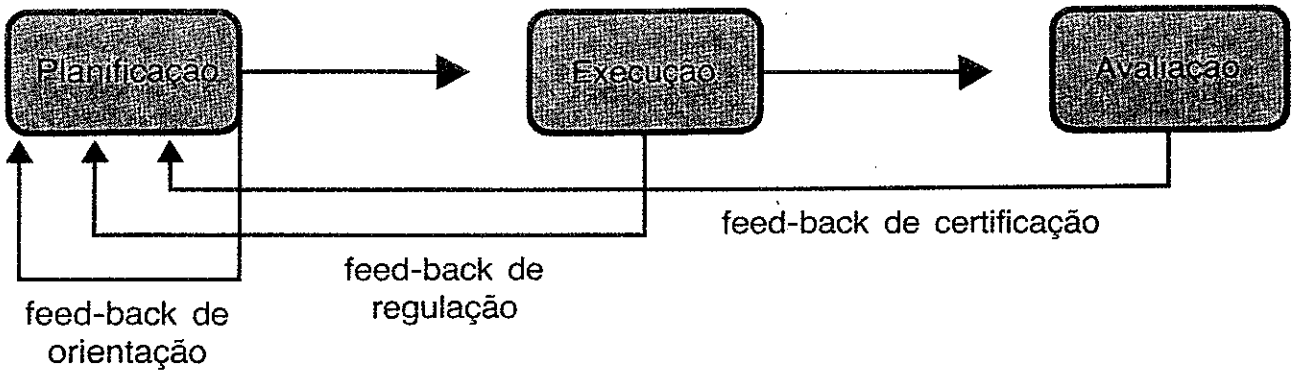


Fig. 1 — *Componentes do sistema ensino-aprendizagem*

(Lemos, V. e Mateus, J., 1980)

A cada uma destas três fases corresponde um tipo de avaliação, consoante a função que têm no sistema (Cardinet, J., 1977). Assim após a planificação (organização das unidades, definição dos objectivos, construção dos instrumentos de avaliação, selecção de métodos e meios) é necessário avaliar em que medida ela está adequada ou não para os alunos que realmente temos. A função é portanto de *orientação* relativamente à implementação da formação. É a avaliação de entrada, de prognóstico ou de diagnóstico.

Durante a fase de execução ou seja, no decurso das actividades de ensino-aprendizagem de uma sequência de ensino, o professor vai analisando com instrumentos objectivos ou subjectivos a forma como está a decorrer a aprendizagem, de modo a obter indicações que lhe permitam manter ou alterar o plano que está a executar, a fim de obter a máxima rentabilidade pedagógica possível. Este tipo de avaliação tem, portanto, uma função de *regulação* do sistema. É a avaliação de regulação, também chamada avaliação formativa (Bloom, 1971).

Finalmente, após o termo de uma sequência de ensino, torna-se necessário verificar em que medida foram atingidos os respectivos objectivos. A função é portanto de *certificação* relativamente à aprendizagem planeada. É a avaliação de certificação, também chamada sumativa (Bloom, 1971).

Devemos fazer notar que nem todos os autores são absolutamente concordantes no que respeita às diversas funções da avaliação, bem como aos diversos momentos de realização.

Assim, a designação «formativa» nem sempre é aceite, por se considerar que qualquer avaliação, em qualquer momento temporal, tem sempre objectivos de formação. Do mesmo modo se passa relativamente à designação de «regulação», considerando-se que qualquer avaliação tem sempre um papel regulador do sistema ensino-aprendizagem.

Outros autores consideram a avaliação de diagnóstico como integrante da avaliação de regulação ou formativa. Outros ainda consideram que não existem três funções, mas quatro, correspondendo a quatro momentos de aplicação, acrescentando às três referidas por nós, a avaliação com funções de diagnóstico que tem por fim determinar as causas de dificuldades reveladas pelo aluno no decurso da aprendizagem (1).

Pensamos, no entanto, que a perspectiva inicialmente exposta (Cardinet, 1977; Lemos e Mateus, 1980; Lemos, 1985) caracteriza as três funções e os três momentos de avaliação, que nos parecem fundamentais.

Avaliação criterial e avaliação normativa

O mais importante em qualquer avaliação de aprendizagem é que ela se baseie numa medição o mais objectiva possível, de todos, de cada um, e só dos objectivos de aprendizagem.

Existem dois tipos fundamentais de análise dos resultados da aprendizagem, que se baseiam em duas concepções diferentes de ensino. Segundo a primeira, designada por «pedagogia tradicional» ou «pedagogia da curva normal», os resultados da aprendizagem dos alunos após um ciclo de ensino-aprendizagem devem distribuir-se segundo a chamada «curva normal» ou «curva de Gauss».

(1) Não se faz aqui uma explanação mais pormenorizada deste assunto por não estar nos objectivos deste trabalho, os quais se situam num plano mais operacional e não na exposição de concepções teóricas diversas.
No entanto, e dado tratar-se de um assunto que nos parece poder vir a merecer do leitor uma análise mais profunda, no final do capítulo são referidas algumas obras que contêm análises mais alargadas do tema.

Nesta curva há uma distribuição simétrica dos resultados relativamente à média fixada, havendo uma concentração de resultados na zona média e uma rarefacção das zonas limites (Fig. 2).

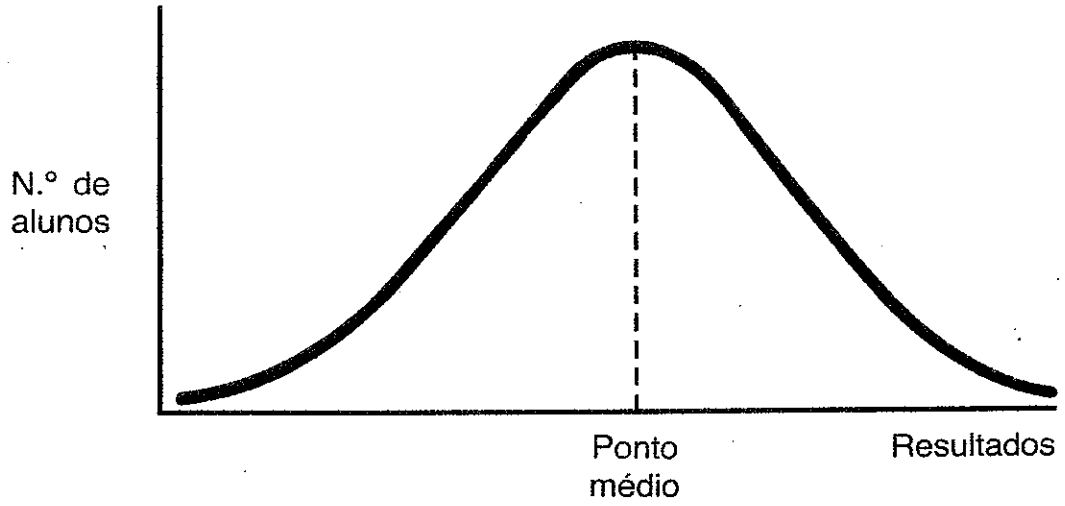


Fig. 2 — *Curva de distribuição normal*

Outra concepção assenta no princípio de que «todos os alunos devem aprender tudo» ou, de uma forma mais realista, que a maioria dos alunos deve atingir a maioria dos objectivos. É a chamada «pedagogia para a mestria» ou pedagogia da curva em «J». Nesta perspectiva os resultados esperados distribuir-se-ão de modo a existir grande concentração no nível máximo e pequena nos níveis abaixo do ponto médio (Fig. 3).

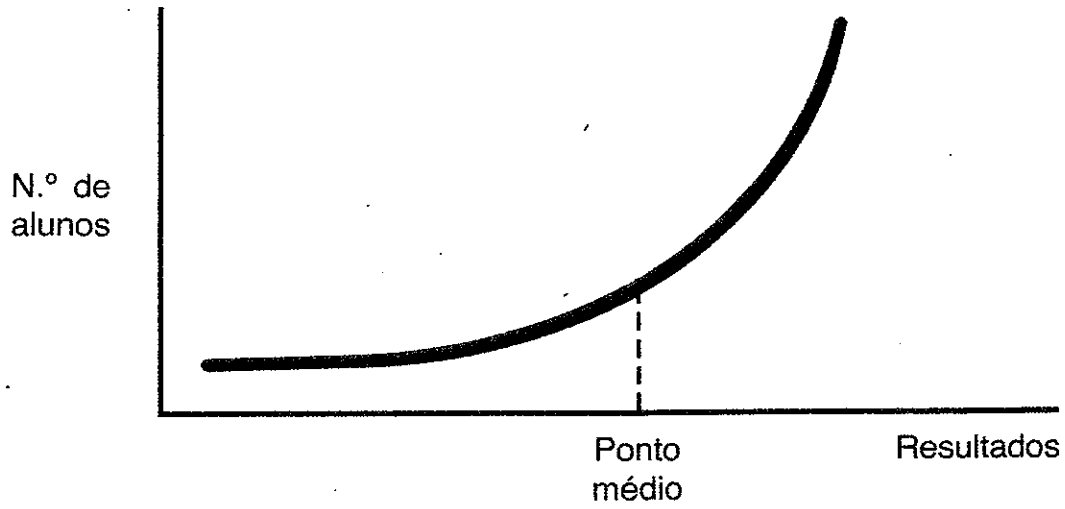


Fig. 3 — *Curva de distribuição em J*

Subjacente a estas duas curvas existem dois conceitos fundamentais de avaliação: a avaliação normativa, em que a avaliação é feita de modo a satisfazer uma distribuição de resultados esperados em curva normal, e portanto, os instrumentos são constituídos de modo a que só metade dos alunos ultrapasse o ponto médio. Não está em causa a consecução de objectivos determinados, mas a posição relativa dos alunos em função do sucesso e insucesso é de 50% para cada um.

Uma outra forma de utilização de curva normal inclui o conceito de desvio-padrão. Trata-se de analisar a distribuição não de forma absoluta em relação à pontuação obtida por cada aluno, mas através de intervalos estatísticos, definidos pelo desvio-padrão (1) (Fig. 4).

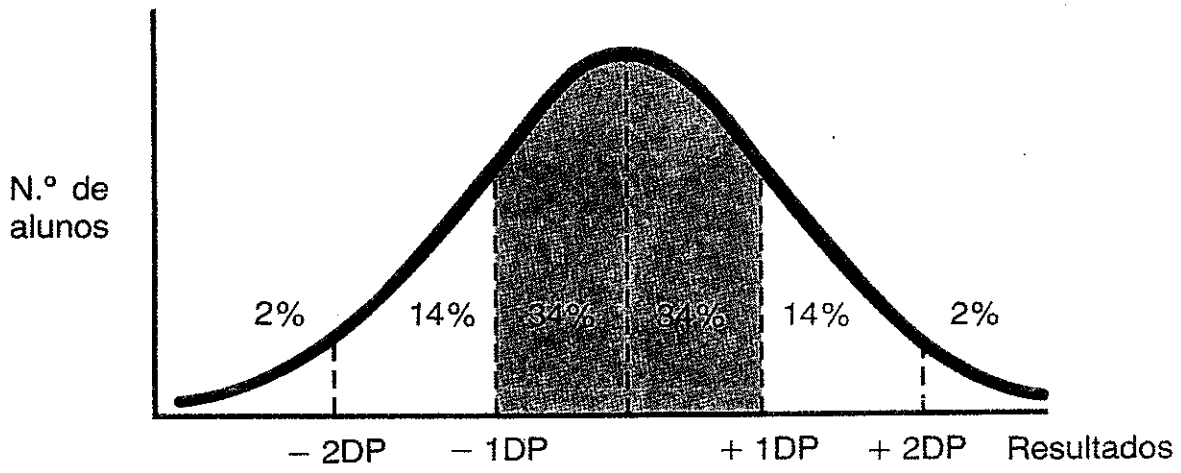


Fig. 4 — *Curva normal com os intervalos determinados pelo desvio-padrão*

Esta forma de análise difere significativamente da anterior no que respeita à concepção de resultados de aprendizagem, mas modifica a actuação em termos de classificação de resultados numa escala, dado que os resultados entre – 1DP e + 1DP se inserem no mesmo nível de escala ou em níveis muito próximos, não sendo, portanto, a separação sucesso/insucesso dependente apenas de um ponto absoluto, mas de um intervalo.

A pedagogia para a mestria assenta não numa comparação de resultados dos alunos em relação ao «normal», mas sim na análise do desempenho relativamente ao critério de sucesso previamente definido. Todos os alunos que satisfaçam o critério têm sucesso e os restantes têm insucesso e quantos mais obtiverem sucesso, mais perfeita é a acção ensino-aprendizagem. Enquanto a perspectiva «normal» assenta

(1) O desvio-padrão é uma medida de variabilidade dentro do grupo. O seu cálculo e utilização estão descritos mais adiante.

numa perspectiva educativa massificada e de selecção relativa, esta assenta numa perspectiva individualizada e de selecção absoluta.

Diversos autores desta concepção consideram que uma acção de ensino tem sucesso, quando 90% dos alunos atingem 90% dos objectivos definidos.

Níveis de aprendizagem

Um aspecto importante a considerar na avaliação é o que podemos designar por «resultados esperados».

Claro que isto faz parte da actividade de planificação. Quando se planifica uma disciplina, área, unidade, sequência, etc., tal faz-se tendo em conta uma população com um determinado nível de pré-requisitos. Após a avaliação de orientação, o plano ajusta-se à realidade encontrada e prevêem-se os «resultados finais». Essa previsão tem um papel importante, porque é ela que vai servir de guia, durante as actividades e de ponto de comparação para os momentos de avaliação de regulação.

Prever resultados, não deve, no entanto, fazer-se no vácuo. Portanto, haverá que definir os chamados níveis de aprendizagem esperada (Gronlund, 1978, Mateus, 1984). Para a definição dos níveis de aprendizagem, haverá que ter em conta um aspecto essencial em todas as actividades de formação. Trata-se do nível mínimo de sucesso. No decurso de uma escala de resultados, *um ponto é na realidade mais importante: o que determina que para um lado está o insucesso e para o outro o sucesso.* Delimita, portanto, o *nível mínimo de aprendizagem.*

Neste nível deve ser incluído o que se considera condição necessária e suficiente para que a aprendizagem possa considerar-se sucedida. Incluem-se aqui os objectivos mais importantes. Não deve, no entanto, confundir-se importância com dificuldade. A importância é determinada pela indispensabilidade de consecução, tendo em conta o ponto de partida (dado pela avaliação de regulação) e o ponto ideal de chegada (determinado no plano).

Para além deste mínimo, a aprendizagem pode (e deve) desenvolver-se para níveis mais elevados. Em termos de facilitação prática, pode considerar-se, portanto, um outro nível de aprendizagem — *o nível de desenvolvimento.* Neste nível estarão, portanto, os objectivos que seria desejável que os alunos atingissem, mas que não é provável que todos atinjam (tendo em conta as diferenças individuais), mas, também não é condição necessária ao sucesso, sendo embora determinante da colocação dos resultados finais numa escala do nível de sucesso.

Em termos práticos e tendo em conta nas etapas de planificação e definição dos níveis de aprendizagem, quais devem ser os resultados esperados quando da planificação de uma unidade de trabalho (após reajustamento proveniente do feed-back produzido pela avaliação de orientação)?

Relativamente ao *nível mínimo* e pela sua própria definição, devemos esperar que a grande maioria dos alunos atinja os objectivos desse nível (90-100% dos alunos para 95-100% dos objectivos) (Fig. 5).

Relativamente ao *nível de desenvolvimento*, e tendo em consideração que, neste caso, as características individuais dos alunos têm muita importância, podemos esperar uma distribuição normal, com desvio à esquerda ou à direita consoante as circunstâncias especificadas de acção. (Fig. 6).

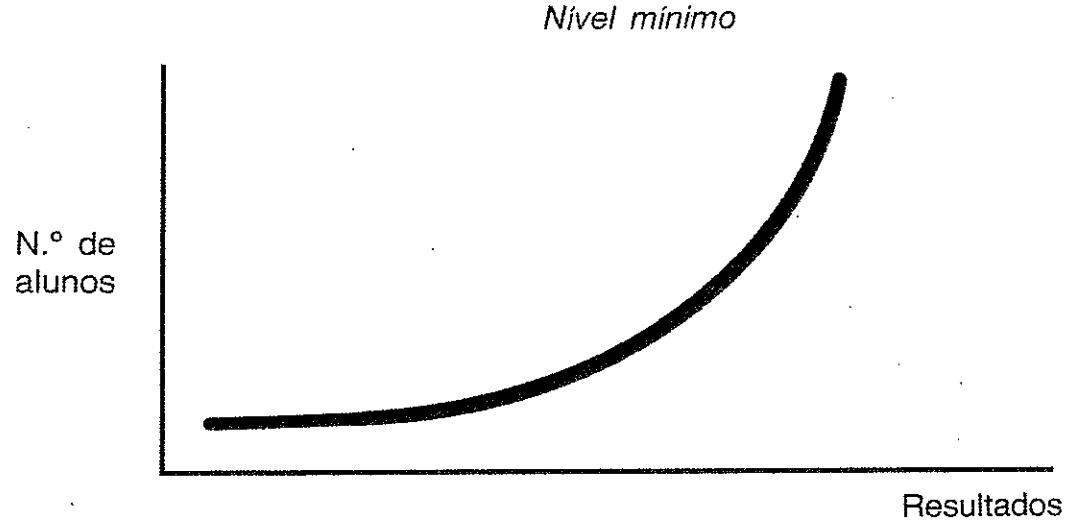


Fig. 5 — *Resultados esperados no nível mínimo*

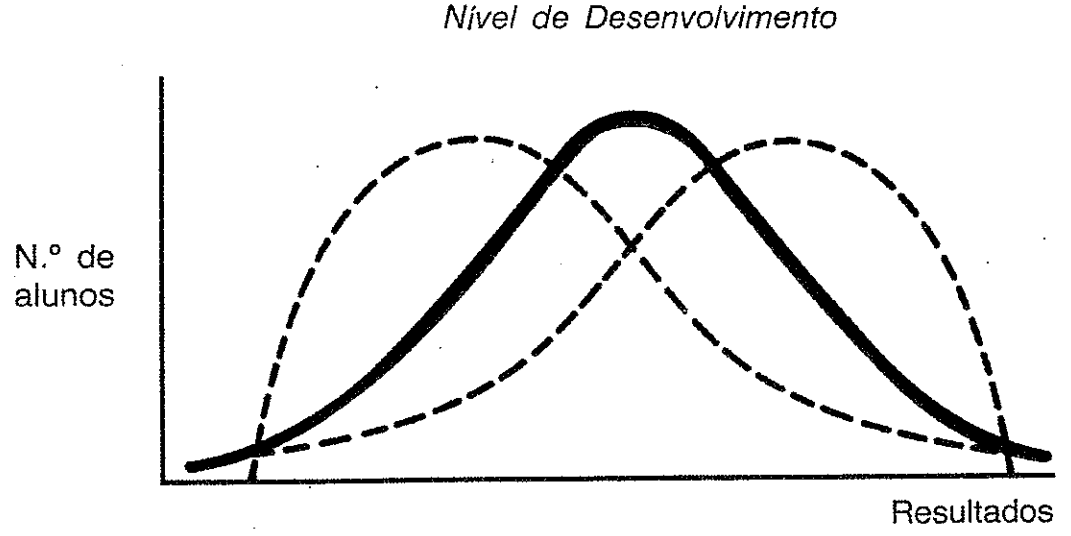


Fig. 6 — *Resultados esperados no nível de desenvolvimento*

Enquadrar estes resultados em escalas de classificação torna-se fácil, dado que a posição dos alunos relativamente ao sucesso não oferece dúvidas, podendo usar-se depois uma metodologia quantitativa ou qualitativa (consoante os casos) para os intervalos de sucesso.

Resumo:

1. A Avaliação tem três funções essenciais:

   Orientação — Diagnosticar a situação dos alunos face aos objectivos planeados para ajustar estes face aos resultados.

   Regulação — Verificar do progresso de aprendizagem em vista aos objectivos fixados, ao longo da execução das diversas partes do plano de ensino.

   Certificação — Verificação do nível final da aprendizagem face aos objectivos definidos, possibilitando o enquadramento numa escala relativa aos níveis de sucesso/insucesso.

2. Basicamente poder-se-ão considerar dois tipos de avaliação, em função dos resultados esperados de uma acção de ensino-aprendizagem.

   Normativa — Quando os níveis de resultados esperados são fixados em termos de distribuição pela população, tomando como padrão a curva normal, não fixando critérios prévios para o sucesso.

   Criterial — Quando são fixados, previamente, os critérios sobre sucesso face a cada um e ao conjunto de objectivos definidos.

3. Na planificação é indispensável a fixação de níveis de aprendizagem necessária:

   Nível mínimo — O conjunto de objectivos cuja consecução é condição necessária e suficiente para o sucesso na acção.

   Nível de desenvolvimento — Inclui os objectivos cuja consecução não é necessária para determinar o sucesso/insucesso, mas é indispensável para colocação na escala final de classificação respeitante aos níveis de sucesso.

BIBLIOGRAFIA:

Birzea, C. (1982). *A Pedagogia do Sucesso*, Lisboa: Livros Horizonte.

Bloom, B.S. et al. (1971). *Handbook on Formative and Summative evaluation of student learning*, N.Y., U.S.A.: McGraw-Hill Book Company.

Bloom, B.S. (1976). *Human characteristics and School learning*, N.Y., U.S.A.: McGraw-Hill Book Co.

Cardinet, J. (1977). *Objectifs educatifs et evaluation individualisée*, Neuchatel, Suisse: Institut Romand de Recherches et Documentation Pedagogiques.

De Ketele, J.M. (1986). A propósito das Noções de Avaliação Formativa, de Avaliação Sumativa, de Individualização e de Diferenciação, in ALLAL, L. et al. (Ed.), *A Avaliação Formativa num Ensino Diferenciado*, Coimbra: Livraria Almedina.

De Landsheere, G. (1971, 1979). *Avaliação Contínua e exames — Noções de Docimologia*, Coimbra: Livraria Almedina.

Gronlund, N. (1978). *Stating objectifs for classroom instruction*, Toronto, Canadá, MacMillan Publ.

Gronlund, N. (1981). *Measurement and Evaluation in Teaching*, 4th ed., N.Y., U.S.A.: MacMillan Publ.

Lemos, V. (1985). *Princípios de Organização Curricular e Avaliação*, Castelo Branco: Escola Superior de Educação.

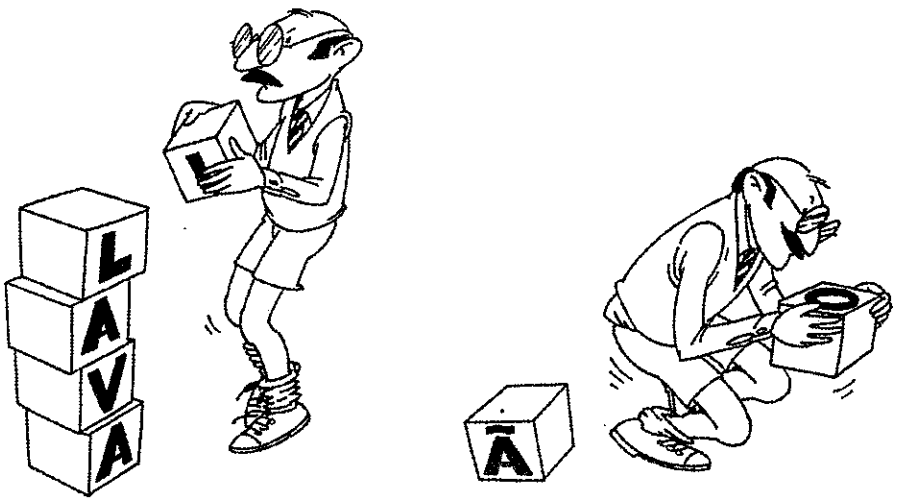
Lemos, V. e **Mateus, J.** (1980). *Os Objectivos em Pedagogia*, Lisboa: Instituto António Aurélio da Costa Ferreira.

Lemos, V. e **Mateus, J.** (1982). *A Avaliação dos Resultados da Aprendizagem*. Sábado (Macau), *21*, 26-29.

Mateus, J. (1984). *O Currículo, a avaliação, a didáctica*, Cacém: Escola Secundária Gama Barros.

2

A CONSTRUÇÃO E A UTILIZAÇÃO DOS INSTRUMENTOS DE AVALIAÇÃO



Existem vários tipos de instrumentos de avaliação da aprendizagem, dos quais consideramos como fundamentais: os testes (objectivos e subjectivos), os registos de observação e os questionários ou inquéritos de opiniões e atitudes.

Cada um dos tipos tem características próprias no respeitante ao que pode medir, à objectividade, ao tempo de execução, à preparação, à correcção, à utilização, etc., pelo que a sua construção obedece a regras diferenciadas, tendo em conta que é indispensável obter sempre a máxima validade, qualidade e funcionalidade, quando se utiliza qualquer instrumento de avaliação.

O quadro I estabelece algumas comparações entre os vários tipos de instrumentos, designadamente no que respeita ao tipo de informação obtida, à objectividade e à relação tempo/resultado.

QUADRO I

TÉCNICAS DE AVALIAÇÃO

	TESTE NÃO OBJECTIVO	TESTE OBJECTIVO	REGISTO DE OBSERVAÇÃO	QUESTIONÁRIO / INQUÉRITO
Tipo de informação obtido	Skills cognitivos Valores Conhecimentos (pouco fiável)	Skills cognitivos Conhecimentos Desempenhos Atitudes	Desempenhos Reacções emocionais Interacção social Skills psicomotores	Opiniões Atitudes Valores Auto-conceito Apreciações subjectivas
Objectividade	Pouco objectivo, sujeito a interferência tendencial (1), sem estabilidade temporal	Objectivo e fiável Estável	Potencialmente objectivo, se os instrumentos forem muito discriminativos, sujeito a interferência tendencial	Pouco objectivo Sujeito a interferência tendencial
Relação Tempo/Resultado	Demorado na análise de resultados. Exige comparações relativas	Pouco demorado na correcção e recolha de resultados. Exige organização e planeamento	Muito demorado. Exige treino de aplicação	Demorado. Exige muita elaboração, quer na preparação quer no tratamento de resultados.

(Adaptado de Ten Brink, T.D., 1982)

(1) No Inglês «bias»

Os testes — tipos e caracterização

Os testes são o instrumento mais usado na medição da aprendizagem. Consistem fundamentalmente numa lista de questões a que os alunos deverão responder. A forma das perguntas e das respostas pode variar, mas o modo como os itens se organizam num teste é muito importante: *deve evitar-se que a estrutura do teste, de algum modo dificulte o trabalho do aluno ou introduza factores exteriores aos objectivos do ensino.* De entre vários aspectos a ter em conta, salientam-se os seguintes:

1 — Cada item deve estar correctamente identificado por um número de série.

2 — O enunciado de um item deve estar completo na mesma página, isto é, o enunciado não deve ser dividido por duas páginas.

3 — O teste deve ser dactilografado ou, no caso de ser manuscrito, deve ser completamente legível em todos os pormenores.

4 — Havendo uma folha de respostas separada, as indicações para a sua utilização devem ser precisas.

5 — A forma de resposta pedida deve ser claramente definida no enunciado.

6 — Quando haja agrupamento de itens, este deve ser organizado em função de objectivos estreitamente correlacionados e, se possível, agrupar itens de forma idêntica.

7 — As indicações sobre a forma das respostas devem ser claramente separadas das componentes da pergunta propriamente dita, através de outro tipo de letra, sublinhado, parêntesis, etc.

8 — Quando existem opções de escolha entre itens, essa opção deve ser feita entre itens de testagem dos mesmos objectivos, e os itens devem ter coeficiente de dificuldade idêntico, além de valor igual em termos de pontuação.

9 — A listagem utilizada deve ser do completo conhecimento e familiariedade dos alunos, nunca introduzindo termos, símbolos ou expressões completamente novas, ainda que aparentemente não pareçam influenciar o texto ou a forma.

10 — O tempo exigido para a realização do teste deve ser calculado em função da velocidade média da realização da turma, adicionado de 20 a 25% de tolerância.

11 — A forma de correcção e pontuação do teste deve ser do prévio conhecimento dos alunos, ou estar claramente definida no próprio teste.

Para além das regras atrás definidas, outras existem no respeitante especificamente à elaboração de cada item em cada teste. Existem dois tipos fundamentais de itens (podendo por vezes constituir um teste inteiro, por si só): os que requerem do aluno uma resposta exacta sem ambiguidades — *testes (itens) objectivos;* e os que

requerem uma resposta que pode variar um intervalo de correcção, não estando sujeita a critérios absolutos exactos — *testes (itens) subjectivos.* No Quadro II estão comparados os dois tipos no que respeita aos aspectos pedagógicos essenciais.

TIPOS DE TESTES — ITENS

	Subjectivo (ensaio)	Objectivo
Competências medidas	Requer expressão do aluno por palavras próprias, usando informação do seu conhecimento. Permite a testagem de operações cognitivas de qualquer nível, bem como a criatividade. Não mede eficientemente a informação factual.	Requer a selecção de respostas correctas entre opções dadas ou de um número limitado de palavras. Permite a testagem de operações cognitivas de qualquer nível, mas não os aspectos criativos. Mede eficientemente o conhecimento de factos.
Tempo de execução	Exige respostas cuja preparação e execução é demorada, o que implica testagens de diversificação limitada.	Exige respostas de execução rápida facilitando a diversificação dos aspectos a testar.
Incentivo aos alunos	Encoraja os alunos a aprender a organizar as suas ideias e a expressá-las. Beneficia os alunos com maior fluência de linguagem escrita, em prejuízo dos restantes.	Encoraja os alunos a estudar e desenvolver competências em toda a área de aprendizagem. Não tem benefícios ou prejuízos discriminados.
Preparação	Exige poucas perguntas por cada testagem. Cada pergunta tem que conter a indicação de limites da resposta.	Cada testagem pode contar muitas perguntas. Exige muita precisão de linguagem, com total ausência de ambiguidades.
Correcção	Exige muito tempo e a utilização de critérios absolutos e relativos e, ainda assim, com falha de precisão.	Exige pouco tempo. É precisa e consistente. Dispensa critérios relativos.

Testes — itens não objectivos

As características a ter em conta para a elaboração de itens de testes não objectivos são as seguintes:

1 — Só devem usar-se para medir objectivos que não sejam mais facilmente mensuráveis por itens objectivos.

2 — Devem cingir-se a problemas específicos, devidamente delimitados na formulação da pergunta.

3 — Não devem exigir respostas demasiado longas, sendo preferível a realização de várias perguntas sequenciais.

4 — *Devem incluir indicações claras acerca do tamanho e profundidade da resposta* (número de razões, exemplos, argumentos, etc.).

Exemplos: (1)

a) "As guerras civis são necessárias para o desenvolvimento social de um país" Defende ou refuta a afirmação anterior, apresentando quatro razões lógicas para os argumentos utilizados e ilustrando cada uma com um exemplo histórico real.

b) Milhões de toneladas de dióxido de carbono pairam sobre o Barreiro, Almada, Seixal e Lisboa. A elas se somam quantidades elevadas de ácidos e óxidos de azoto, de chumbo e de cádmio. Verifica-se ao mesmo tempo que a densidade populacional tem vindo a crescer nestas regiões.
Estabelece uma relação entre os dois fenómenos, referindo duas causas possíveis para os mesmos.

Testes — itens objectivos

Existem vários tipos de itens objectivos, de acordo com a forma da resposta exigida:

Resposta curta — são os que exigem uma resposta limitada a uma palavra ou frase por inteiro, ou a completar frases, textos, diagramas, esquemas, etc., previamente fornecidos. Os aspectos a ter em conta na sua construção são os seguintes:

1 — A resposta correcta deve poder ser dada, sem qualquer ambiguidade numa palavra ou frase.

(1) Nota do Autor — Os exemplos de itens usados em todo o livro, foram já usados, tenso sido retirados de testes produzidos e utilizados por professores na sua actividade, pelo que agradecemos a todos os colegas que nos cederam os respectivos testes.

2 – A questão deve incluir as indicações sobre a linguagem a usar na resposta (palavras, expressões numéricas, aritméticas ou geométricas, símbolos, unidades de medida, fórmulas, etc.)
3 – Quando se trata de completar esquemas, frases, legendas, expressões matemáticas, etc., deve pedir-se um número restrito por item.
4 – Partes a completar nunca devem ser colocadas no início da frase ou expressão.

Exemplos:

a) Qual o nome que se dá a uma estrutura composta por:
1 – vários menhires em linha?
2 – vários menhires em círculo?

b) 1. Completa a expressão seguinte:
$9x^2 - 25 = (3x - 5)$ ( )
2. Considera agora que:
$9x^2 - 25 = 0$
Calcula o valor de x, registando as operações de cálculo.

c) "O seu rasto estendia-se pelo céu nocturno, com o fulgor de gigante chama".
1. Na frase anterior é utilizada a palavra "chama". Indica dois sinónimos desta palavra, de acordo com o sentido da frase.
2. A palavra anterior pode também ser usada como uma forma do verbo "chamar". Indica dois substantivos derivados deste verbo.

Verdadeiro/Falso — exigem uma resposta identificativa da veracidade de afirmações dadas. Características para elaboração:

1 – Cada afirmação ou expressão deve ser de significado definido e sem qualquer ambiguidade.
2 – Cada afirmação ou expressão deve ser verdadeira ou falsa em absoluto, sem qualificação ou excepções.
3 – Cada expressão deve ser curta e directa, evitando considerandos ou qualificações.
4 – A expressão deve ter uma só interpretação possível, evitando, portanto, as de "duplo sentido".
5 – Só devem ser usadas expressões afirmativas, evitando sempre o uso de expressões definidas pela negativa.
6 – Não devem conter advérbios de modo tais como geralmente, sempre, nunca, por vezes, etc.
7 – Tratando-se de tornar verdadeira uma afirmação falsa, deve-se exigir somente a modificação de uma ou duas palavras ou símbolos e nunca toda a expressão ou a sua estrutura.

Exemplos:

a) Diz quais das seguintes frases são falsas e quais são verdadeiras:
   1 – Uma comunidade é composta pelo conjunto dos seres vivos de um ecossistema.
   2 – O conjunto dos seres vivos da mesma espécie, num determinado ecossistema designa-se por população.
   3 – Os seres inanimados de uma comunidade são o seu ecossistema.
   4 – Num ecossistema podem existir várias populações.

b) Das expressões seguintes indica as verdadeiras:

1. $\frac{3}{4} = \frac{9}{12}$
2. $0{,}625 = \frac{8}{5}$
3. $3 = \frac{27}{9}$
4. $\frac{3{,}5}{7} = \frac{2}{4}$

Combinação — exigem que seja estabelecida uma relação entre pares de palavras, frases ou símbolos. Características para a construção:

1 – Cada par de listas deve reportar-se somente a um objectivo ou conceito.
2 – Uma das listas deve ser cerca de 50% maior que a outra, e a mais pequena não deve conter mais que seis ou sete itens.
3 – A cada premissa de uma lista deve corresponder, sempre que possível, uma só resposta na outra lista.
4 – Quando seja exigida uma combinação entre uma premissa numa lista e várias na outra, deve ser indicado o número destas que é requerido.
5 – Não devem usar-se expressões de "duplo sentido". As expressões devem ser claras e objectivas.

Exemplos:

a) Relaciona *cada uma* das designações da coluna da esquerda com *uma* das designações da coluna da direita.

1. Rocha metamórfica	*a*) Granito
2. Rocha sedimentar	*b*) Feldspato
3. Rocha eruptiva	*c*) Xisto
	d) Arenito

b) *Write 5 sentences*, relating column A with column B:

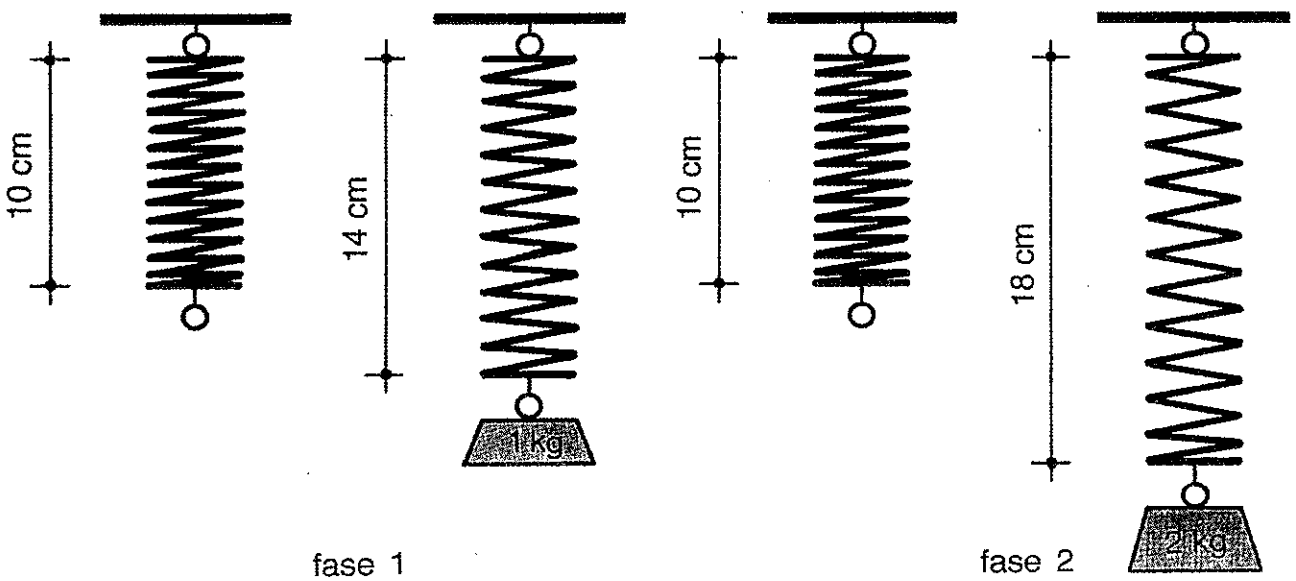
A	B
1. I have	*a*) a teacher
2. I am	*b*) Kind
3. I go	*c*) to London
	d) book
	e) one computer
	f) to school

Escolha múltipla — Exige a selecção de uma resposta correcta de entre várias opções dadas. Características para construção:

1 – Referir-se a um só problema ou questão.
2 – Cada lista de opções deve conter só uma resposta correcta.
3 – Não devem usar-se expressões definidas pelo sentido negativo.
4 – As alternativas apresentadas devem ser similares, em palavras, estilo, tamanho, etc.
5 – As alternativas incorrectas devem ser plausíveis, tanto quanto possível.
6 – As alternativas devem ser mutuamente exclusivas.
7 – Não devem usar-se expressões de "duplo sentido".
8 – Deve evitar-se o uso de alternativas "nenhuma das anteriores".
9 – O número ideal de opções é de quatro e não deve ser inferior a três.
10 – As palavras que se devem repetir em todas as opções devem fazer parte da base da pergunta e não das opções.

Exemplos:

A figura seguinte representa fases sucessivas de uma experiência, em que se suspenderam de uma mola elástica, vários corpos.



Escolhe das afirmações seguintes aquela que consideras que pode corresponder a uma conclusão válida da experiência efectuada.

1. Os comprimentos da mola são directamente proporcionais às intensidades das forças nela aplicadas, dentro do limite de elasticidade da mola.
2. **Os alongamentos da mola são directamente proporcionais às intensidades das forças nela aplicadas, dentro do limite de elasticidade da mola.**
3. Os comprimentos da mola são inversamente proporcionais às intensidades das forças nela aplicadas, dentro do limite de elasticidade da mola.
4. Os alongamentos são inversamente proporcionais às intensidades das forças aplicadas à mola, dentro do limite de elasticidade da mola.

Registos de observação

Estes instrumentos podem usar-se para verificação da frequência de determinados comportamentos e qualidade dos desempenhos específicos. Podem assumir basicamente duas formas: lista de verificação ou grelha de análise.

Listas de verificação

As listas de verificação não se usam geralmente quando se trata de frequência de comportamentos, mas somente quando se trata da simples presença ou ausência dos mesmos. Para elaboração de uma lista de verificação, é necessário identificar previamente os comportamentos ou características importantes. As propriedades de uma boa lista de verificação, são as seguintes:

1 – Deve ser curta.
2 – Cada item deve ser claro e o mais objectivo possível.
3 – Cada item deve referir-se a uma só característica ou comportamento.
4 – Devem estar incluídas características ou comportamentos importantes.
5 – Deve ser facilmente manuseável (por exemplo, nunca estar escrita na frente e costas de uma folha).

Exemplo:
Avaliação de uma pintura baseada em figuras geométricas.
— Estão presentes todas as figuras pedidas.
— Não denota erros na construção das figuras.
— Denota equilíbrio na distribuição das figuras de linhas rectas e de linhas curvas.
— Não deixou visíveis traços auxiliares de construção.
— Não usou cores primárias (tal como era indicado).
— A execução das figuras é rigorosa.

— A execução da pintura é rigorosa, não havendo ultrapassagens ou sobreposições indevidas.
— A combinação das cores é agradável.
— Cada tonalidade apresenta aspecto uniforme.
— Utilizou todo e só o espaço permitido.

Grelhas de análise

Devem utilizar-se quando se pretende observar a frequência de determinados comportamentos. Para elaboração de uma grelha de análise é necessário seleccionar previamente os comportamentos cuja frequência se deseja observar. As propriedades de uma grelha de análise são as seguintes:

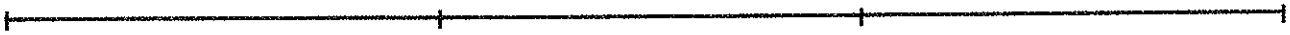
1 – Deve referir-se a um número reduzido de comportamentos.
2 – Deve estar organizada de modo a que seja possível assinalar cada ocorrência de forma simples, fácil e rápida.
3 – Deve ser facilmente manuseável (nunca frente e costas de uma folha).

Exemplo:

Comportamentos / Datas	Interrompe o professor	Interrompe os colegas	Ajuda os colegas	Arruma o material	Não cumpre as tarefas em grupo
1.ª Semana					
2.ª Semana					
3.ª Semana					
TOTAIS					

As grelhas de análise podem ser complementadas por *escalas qualitativas*, cuja função é a de avaliar a qualidade de um comportamento ou característica. Um exemplo de um item de uma escala qualitativa é o seguinte:

Em que medida o coordenador do grupo encoraja a discussão?



Desencoraja com comentários negativos	Não encoraja nem desencoraja	Encoraja com comentários positivos

Embora possam ser instrumentos úteis, a sua utilização pelo professor torna-se geralmente difícil, pelo que devem usar-se somente quando a qualidade de um comportamento, não observável de outro modo, seja realmente importante.

Questionários de opinião

Os questionários utilizam-se para obter informações sobre as opiniões e apreciações dos alunos, de modo a poder avaliar o nível de desenvolvimento de determinados valores e atitudes. Para construir um questionário devem ter-se em conta os seguintes aspectos:

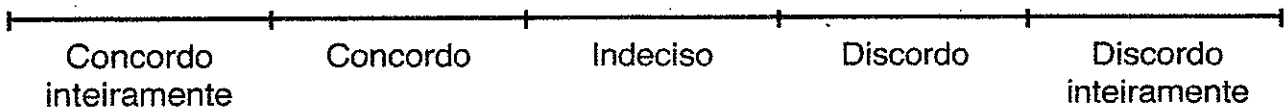
1 – Definir previamente quais os valores ou atitudes que se pretendem medir.
2 – Seleccionar a forma adequada (existem várias) das perguntas.
3 – Seleccionar a forma (existem várias) em que as respostas devem ser dadas.

Além destes aspectos prévios de construção, um bom questionário tem as seguintes propriedades:

1 – Não deve ser demasiado longo para evitar o desinteresse a a fadiga.
2 – Deve preferencialmente exigir a mesma forma de resposta a todas as questões. Quando seja necessário mais de uma forma, os itens devem ser agrupados em função do formato de resposta exigido.
3 – As questões ou afirmações devem ser claras e sem expressões de «duplo sentido».
4 – Conforme o formato das questões devem estas seguir as regras respectivas a esse formato.

As formas possíveis para os itens de questionários enquadram-se nas formas já analisadas para testes e registos de observação. Assim podem ser:

1 – Resposta curta.
2 – Escolha múltipla
3 – Escalas — uma das escalas mais usadas em questionários é a escala de Lickert em que, face a uma afirmação, o aluno assinala um dos 5 graus:



Outro tipo de escalas são as de ordenação, em que o aluno tem de ordenar as afirmações fornecidas da primeira à última.

Exemplo:
Na aula prefiro:
— ouvir o professor
— ouvir os colegas
— falar com os colegas
— falar com o professor
— trabalhar sozinho
— trabalhar em grupo

Podem usar-se também o tipo de escalas já referidas para os registos de observação, ou ainda as que exigem respostas: sim/não, verdadeiro/falso, etc.

BIBLIOGRAFIA:

De Ketele, J. (1982). *Docimologie — Introduction aux concepts et aux pratiques*, Louvain-la--neuve, Belgique: Cabay Ed.

Ebel, R. e **Frisbie, D.** (1986). *Essentials of Educational Measurement*, 4th Ed., Englewood Cliffs, N.Y., U.S.A.: Prentice-Hall, Inc.

Gronlund, N. (1979). *Elaboração de Testes para o Ensino*, S. Paulo, Brasil: Livraria Pioneira Editora.

Lemos, V. (1985). *Princípios de Organização Curricular e Avaliação*, Castelo Branco: Escola Superior de Educação.

Mateus, J. (Coordenação) (1985). *Formação por Competências — Relatório 1984/85*, Lisboa: Projecto FOCO — Faculdade de Psicologia e Ciências da Educação.

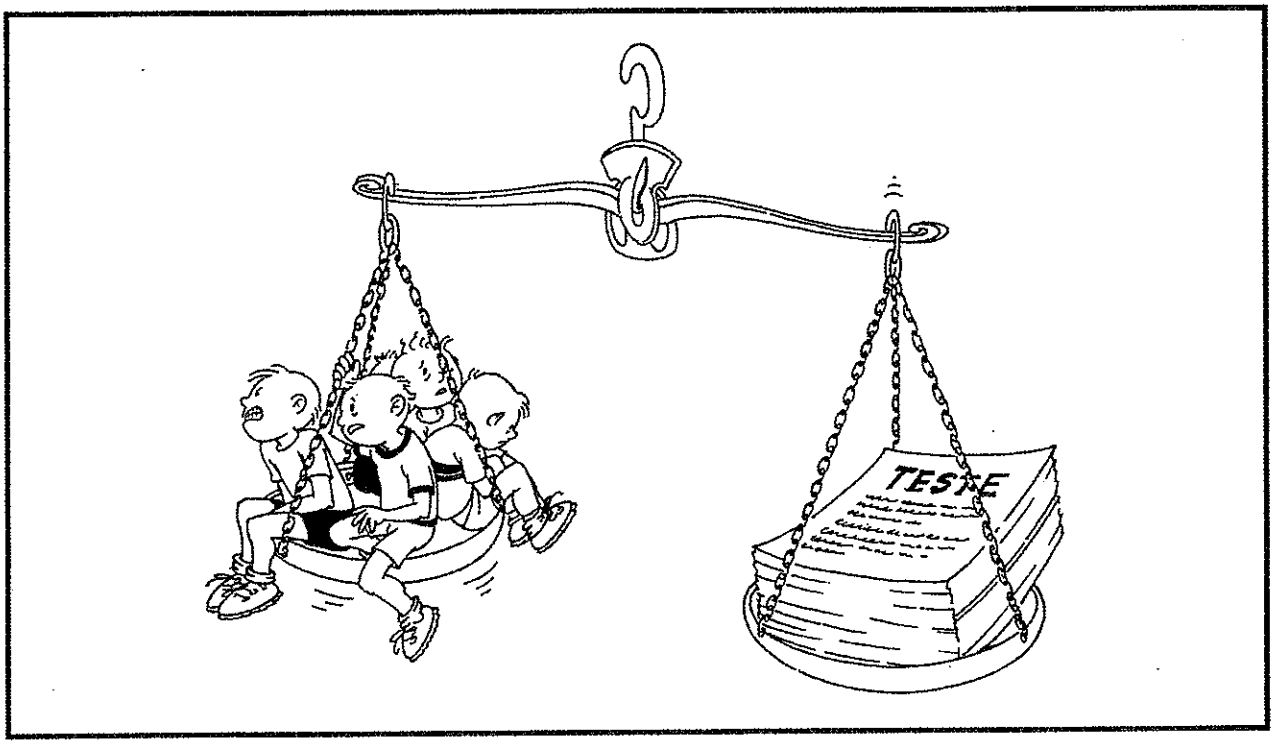
Mateus, J. e **Lemos, V.** (1982). Os Instrumentos de Avaliação da Aprendizagem, *Sábado* (Macau), *23*, 22-25.

Ten Brink, T.D. (1982). Evaluation. In COOPER, J. (ed.), *Classroom Teaching Skills*, 2nd ed., N.Y., U.S.A.: DC Heath and Company.

Tinkelman, S. (1971). Checklist for Reviewing Local School Tests, in MCFARLAND, S. e HEREFORD, C. (ed.), *Statistics and Measurements in the classroom*, Dubuque, Iowa, U.S.A.: WM.C. Brown Company Publ.

3

QUALIDADE DOS TESTES



A qualidade dos testes afere-se por várias características, as quais podem ser ordenadas e agrupadas de diversos modos. Pensamos, que as quatro seguintes são, na realidade as mais importantes e significativas:

Validade — O teste deve requerer do aluno exactamente os mesmos comportamentos (acções, operações, conhecimentos) definidos nos objectivos, e sob idênticas condições.

Diferencialidade — As operações (acções, conhecimentos) requeridas só devem poder ser realizadas pelos alunos que alcançarem os objectivos.

Fidelidade — O teste deve constituir uma medida consistente da consecução dos objectivos, em qualquer momento ou circunstância, estando portanto liberto de factores estranhos que possam influenciar o desempenho do aluno.

Objectividade — Não deve haver dúvidas quanto ao critério de sucesso do aluno na resolução do teste. Para testes não objectivos deve usar-se o sistema de concordância de dois ou mais observadores, além de um procedimento de correcção baseado em critério relativo.

Validade

As questões sobre a validade dos testes e outros instrumentos de avaliação, têm sido discutidas bastantes vezes e não são de fácil resolução. Conforme se disse na Introdução, na maioria das obras sobre o assunto, os procedimentos de validação referem-se quase sempre a testes psicológicos padronizados. Embora tais aspectos também possam aplicar-se para a elaboração de testes de aprendizagem de âmbito alargado, tais como por exemplo, testes de exame ou aferição de carácter nacional, a verdade é que no dia-a-dia do professor, na testagem de unidades ou sequências de ensino, tais procedimentos não são possíveis, ou sendo-o, a maioria das vezes são impraticáveis, tendo em conta o tempo de execução e outros requisitos, que o professor em actividade normal, não tem possibilidades de cumprir.

De qualquer modo, será conveniente esclarecer o conceito de validade, antes de analisar procedimentos. A validade diz respeito *ao que o teste mede*. Refere-se portanto à medida da certeza de que *um teste mede realmente aquilo que se pretende medir*, ou, de outro modo, à medida da relação entre as questões do teste e os objectivos que se pretendem medir.

Distinguem-se, normalmente, quatro tipos de validade: conteúdo, conjunto, concorrente e preditiva.

As validades de conjunto, concorrente e preditiva respeitam fundamentalmente a testes psicológicos. Assim a validade de conjunto refere-se à ligação e afinidade existente entre os comportamentos testados. É uma característica relativa aos testes de inteligência e personalidade. Por exemplo num teste de medição da atenção, o conjunto das capacidades medidas em todo o teste deve representar efectivamente a capacidade geral de atenção.

A validade concorrente refere-se à relação com outro instrumento de medida. Por exemplo, quando um investigador desenvolve um teste de raciocínio (ou outro tipo), os resultados obtidos neste teste, pelos testados, devem estar estreitamente correlacionados com os resultados obtidos noutra forma de medida já existente da mesma capacidade.

A validade preditiva refere-se à correlação entre o resultado obtido no teste e os comportamentos futuros do indivíduo testado. Respeita fundamentalmente aos testes de aptidão. Por exemplo, a correlação entre os resultados obtidos num teste de aptidão para vendas e os resultados obtidos pelo indivíduo nessa actividade, estabelece o nível de validade preditiva do teste.

Estes três tipos de validade têm pouco interesse pedagógico, pois referem-se à validade da medição de características ou aptidões psicológicas e não de aprendizagens específicas, sequenciais e construtivas. A validade de conjunto tem mesmo interesse exclusivamente psicológico (Cronbach e Meehl, 1955). A validade concorrente pode, em avaliação escolar, ser utilizada para formas diferentes de testagem dos mesmos objectivos, mas colocando sempre o problema de qual a forma que se está a tentar validar. Por sua vez a validade preditiva pode ter-se em conta, no respeitante aos instrumentos de avaliação de regulação (formativa), (relativamente à avaliação de certificação), mas, tendo em conta os procedimentos exigidos (calcula-se normalmente através do teste de Pearson), torna-se pouco praticável no dia-a-dia.

Portanto, diremos, como grande parte dos autores (Lindeman, 1967; Goring, 1981 e Landsheere, 1979) que a validade de conteúdo é a que realmente mais interessa ao professor nas suas avaliações de aprendizagem.

A validade de conteúdo

A validade de conteúdo refere-se à relação entre as perguntas e os objectivos fixados.

Não se trata de uma medida estatística, embora um tratamento estatístico referente à análise de itens possa ser usado para contribuir, à posteriori, para a avaliação de validade.

Trata-se mais de um procedimento de rigor na construção do teste, de modo a que ele se refira todo e só aos objectivos de aprendizagem fixados e estudados, reflectindo também (na quantidade de itens ou na forma de pontuação) a importância relativa de cada objectivo, a qual está directamente relacionada com o ênfase e o tempo dispendido nas actividades de ensino-aprendizagem respectivas.

Algumas regras elementares podem ser enumeradas:

1 — Todos os objectivos de aprendizagem (em cada unidade ou sequência) que foram estudados, devem ser testados.
2 — Cada objectivo deve ser testado separadamente no contexto do teste.
3 — Cada item deve medir somente um e um só objectivo.

4 – O número de questões por objectivo deve ser o necessário e suficiente para assegurar da consecução desse objectivo.
5 – Havendo pontuação diferenciada, ela deve corresponder à importância (ênfase no ensino-aprendizagem) dos objectivos.

Juntando a estas instruções as que foram definidas para a construção dos testes, poderá existir uma relativa margem de segurança no respeitante à validade do teste que se está a utilizar.

Existem no entanto mais alguns procedimentos que podem obviar às ameaças à validade. Um deles, respeita à apreciação do teste por outros colegas. Face à lista de objectivos e ao teste, se dois professores da mesma área ou disciplina, discordarem quanto à validade da medição de algum objectivo por um dos itens, este deve ser imediatamente reformulado ou substituído por outro que não deixe margem a dúvidas.

Um outro aspecto diz respeito às opiniões, que os próprios alunos emitem sobre o teste. Quando um professor ouvir dos seus alunos o comentário: "Não foi isto que «demos» na aula"; em vez de tentar justificar que a afirmação está errada ou incorrecta (atitude muito comum) deve sim, reanalisar imediatamente o teste, porque normalmente, trata-se de uma questão de validade, em geral relacionada com uma relação indevida entre a importância (ênfase) do(s) objectivo(s) e o respectivo peso no teste, quando não se trata de um item inadequado ao objectivo, o que sucede com frequência maior que a desejável.

Diferencialidade

Após a aplicação do teste e da sua correcção deverá, sempre que possível, ser realizado um trabalho que designamos por análise de itens e que consiste em determinar o índice de discriminação, e o grau de dificuldade, bem como a análise dos erros e omissões dos alunos. Trata-se portanto, de determinar as características de diferencialidade do teste.

Grau de dificuldade

Começa-se por considerar os resultados hierarquizados, retirando-se em seguida os testes — respostas respeitantes aos 27% piores resultados e igualmente aos 27% melhores resultados (1). Partindo destes testes construir-se-á um quadro semelhante ao seguinte:

(1) Tendo em conta o tamanho normal das turmas nas nossas escolas, poder-se-á considerar 25% em vez de 27%, o que torna o cálculo mais fácil e mais rápido, sem ameaçar a segurança estatística.

ITEM		Grupo pior	Grupo melhor	TOTAL
1	C	5	8	13
	E	5	2	7
2	C	8	9	17
	E	2	1	3
3	C	7	6	13
	E	3	4	7

Legenda: Grupo pior — grupo com resultados mais baixos
Grupo melhor — grupo com resultados mais altos
C — resposta correcta
E — resposta errada

Partindo destes dados é possível calcular o índice de dificuldade e o de discriminação de cada item.

Assim para o cálculo da dificuldade:

$$Df = \frac{M + P}{N}$$ sendo

Df – grau de dificuldade
N – número total de alunos de ambos os grupos
M – número de alunos do grupo melhor que responderam erradamente
P – número de alunos do grupo pior que responderam erradamente

Para o exemplo indicado, relativamente ao item 1 teríamos:

$$Df = \frac{2 + 5}{20} = \frac{7}{20} = 0{,}35$$

Índice de discriminação

O índice de discriminação, indica até que ponto cada pergunta discrimina entre os níveis de competência dos alunos. Parte-se do princípio que será de obter maior percentagem de alunos a responder correcto naquele item entre o grupo que obteve melhores resultados globais do que entre o grupo que obteve piores resultados.

Calcula-se o índice de discriminação de acordo com a fórmula seguinte:

$$Dc = \frac{M - P}{N/2}$$ sendo

Dc – índice de discriminação
N – número total de alunos de ambos os grupos
M – número de alunos do grupo melhor que respondem correctamente
P – número de alunos do grupo pior que respondem correctamente

Para o exemplo dado atrás, teríamos:

$$Dc = \frac{8 - 5}{20/2} = \frac{3}{10} = 0{,}30$$

Utilização dos índices de dificuldade e discriminação

O índice de discriminação poderá, portanto variar entre + 1,00 e – 1,00, e o índice de dificuldade entre 0,00 e + 1,00.

Após esta breve análise dos procedimentos matemáticos, poder-se-á perguntar qual a utilidade do cálculo de índices e quais os níveis desejáveis e adequados para cada índice?

A utilidade é clara, pois permitirá eliminar perguntas que não demonstrem a validade estatística necessária e permitirá também identificar os itens que demonstram boa validade, podendo usá-los mais tarde com alguma segurança (embora a validade respeite ao grupo a que se aplicou o teste, o professor trabalha muitas vezes com grupos semelhantes em condições idênticas). Poderá o professor, ainda, ir construindo uma bateria de objectivos/itens relativamente segura, que poderá usar noutras oportunidades.

Quanto ao nível desejável ou adequado dos índices para inferir da validade, haverá que separar o que respeita ao *nível mínimo de aprendizagem*, e ao *nível de desenvolvimento* (como se falou atrás).

Evidentemente que com um procedimento como o descrito, os índices para os dois níveis serão necessariamente diferentes, pois que o primeiro refere-se ao que *todos* os alunos deveriam aprender.

Relativamente ao nível mínimo dir-se-á que não é lógico considerar níveis mínimos de dificuldade e discriminação dos itens, porque a resposta correcta por parte de quase todos os alunos, é, ela mesmo, o resultado esperado, podendo portanto o grau de dificuldade e o índice de discriminação aproximarem-se de zero, embora seja desejável que não alcance sempre este valor. Para o nível mínimo, portanto, estes não são procedimentos adequados para determinar a validade.

Relativamente ao nível de desenvolvimento, a situação é bem diferente, sendo estes índices adequados para ajudar à determinação da validade.

Assim quanto aos níveis desejados Ebel (1986) refere que para testes de escolaridade um índice de discriminação abaixo de 0,19 deve implicar a rejeição do item, podendo considerar-se bom acima de 0,40, razoável de 0,30 a 0,39 e necessitar de reformulação quando for de 0,20 a 0,29.

Relativamente à dificuldade, a maioria dos autores considera 0,30 a 0,70 como o intervalo desejável, com tolerância de ± 0,10, consoante as circunstâncias.

Deverá ainda referir-se que um índice de discriminação negativa, deverá implicar, em qualquer situação, a rejeição imediata do item.

A análise de itens fornece ainda outros dados que devem ser usados para apreciar a validade do teste e dos itens e ainda para melhorar a formulação dos itens que se revelem deficientes.

Assim, por exemplo, relativamente a itens de escolha múltipla, de correlação, verdadeiro-falso, etc., poderá elaborar-se um quadro em que para cada item se assinala a opção, ou correlação, ou identificação feitas pelos alunos.

Partido destes dados pode verificar-se se um item não foi respondido pela maioria dos alunos, quais as opções mais assinaladas pelos alunos, quais as respostas falsas mais escolhidas, etc., e assim avaliar da plausibilidade das diversas opções de resposta, da falta de objectividade de formulação, de factores subjacentes de confusão conceptual, etc.

Do mesmo modo convém fazer um quadro em que para cada aluno é assinalada a posição para cada item, pois a partir dele é que se verifica a situação de cada aluno relativamente a cada objectivo, permitindo reajustar quer planos globais de revisão para toda a turma, quer actividades individuais ou grupais com fins de remediação ou enriquecimento.

Fidelidade

A fidelidade refere-se à característica de um teste *medir com exactidão* o que se pretende. Na actividade do professor esta característica depende na sua maior parte dos procedimentos já indicados para melhorar a validade, bem como os índices relativamente à construção.

Convirá dizer que esta característica está também mais ligada a testes de aplicação geral numa população, relativamente à medida de uma característica por todo o teste. No caso da actividade pedagógica, em que para cada objectivo de aprendizagem é construído um item (ou mais) e o total desses itens é que forma o teste, esta questão tem pouca importância, dado que o mais importante é cada item de «per si», tendo em conta que o teste inclui os itens referentes a todos os objectivos.

No entanto, quando o professor não define os objectivos e constrói para cada um deles um item, e faz uma selecção aleatória (ou não) de conceitos ou princípios no âmbito da «matéria ensinada» ou ainda resolve só testar (ou retestar) alguns dos objectivos definidos, deve, nestes casos, fazer o estudo da fidelidade do teste.

Para comprovar a fidelidade existem três tipos de procedimentos fundamentais:

a) Teste — Reteste — Aplica-se o teste novamente após algum tempo de se ter aplicado pela primeira vez. Calcula-se depois o coeficiente de correlação dos resultados(1).

b) Formas equivalentes — Aplicam-se duas formas consideradas equivalentes (extensão, dificuldade e conteúdo) do mesmo teste, calculando-se a seguir o coeficiente de correlação dos resultados.

c) Método das metades (split-half) — Divide-se o teste ao meio (por exemplo, perguntas pares e perguntas ímpares). Calcula-se depois o coeficiente de correlação dos resultados obtidos em cada metade.

Dos métodos indicados é evidente que o primeiro é de pouca usabilidade pelos professores, sendo mais adequados a outros dois. A prática de aplicar duas formas do mesmo teste é, aliás, frequente em muitos professores, designadamente para obviar a possíveis comportamentos de «cópia» pelos alunos. O método das metades, sendo menos usado, é, no entanto, o mais adequado para os professores em geral, que aplicam somente uma forma do teste, não ficando por tal impedidos de calcular a fidelidade, quando seja caso disso, como já foi referido.

Objectividade

O grau de objectividade não é uma medida numérica, mas qualitativa, comprovando-se com os procedimentos utilizados para a construção e aplicação do teste.

Os procedimentos para garantir a objectividade estão descritos atrás, quando se analisaram as características dos testes objectivos.

A utilização de testes não objectivos, não implica no entanto a sua completa subjectividade, a qual pode ser reduzida, quando se têm em conta procedimentos adequados (ver regras para a elaboração de testes-itens não objectivos). A forma de obviar à subjectividade na correcção e pontuação será analisada mais adiante.

(1) O coeficiente de correlação pode calcular-se utilizando procedimentos de estatística inferencial como o teste de Pearson, o teste de Spearman-Brown ou de Kuder-Richardson.
Neste trabalho não vamos analisar procedimentos de estatística inferencial por estarem fora do seu âmbito. No entanto a descrição de tais procedimentos pode ser encontrada em qualquer das obras sobre estatística indicadas no final deste capítulo e do seguinte.

BIBLIOGRAFIA:

Davis, R. et. al. (1979). *Sistemas de Aprendizagem*, 2.ª ed., S. Paulo, Brasil: McGraw-Hill do Brasil, Ltda.

Diederich, P. (1971). Short-Cut Statistics for Teacher — Made Tests, in McFARLAND, S. e HEREFORD, C. (ed.), *Statistics and Measurement in the Classroom*, Dubuque, Iowa, U.S.A.: WM. C. Brown Company Publ.

Ebel, R. e **Frisbie, D.** (1986). *Essentials of Educational Measurement*, 4th ed., Englewood Cliffs, N.Y., U.S.A.: Prentice-Hall Inc.

Lindeman, R. (1978). *Medidas Educacionais*, Porto Alegre, Brasil: Editora Globo.

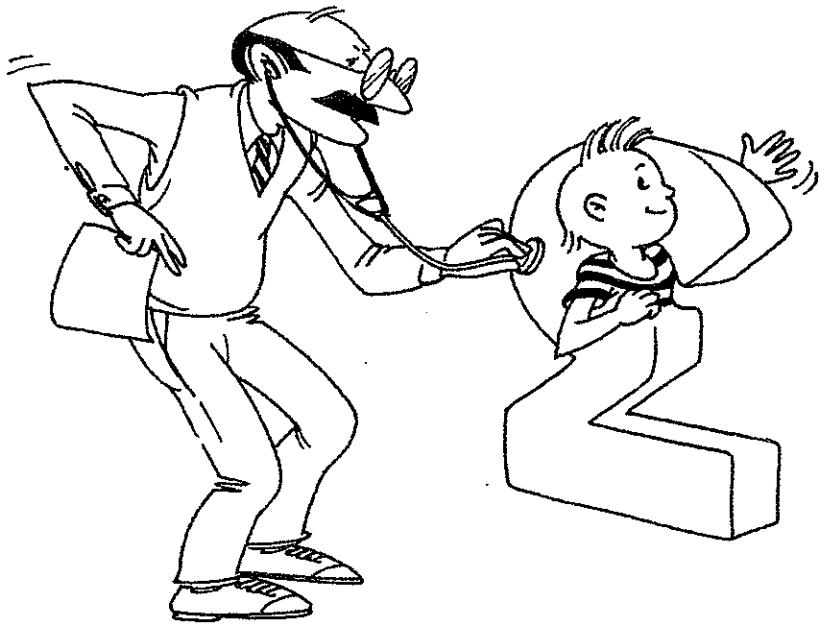
Vallejo, P. (1979). *Manual da Avaliação Escolar*, Coimbra: Livraria Almedina.

Vianna, H. (1982). *Testes em Educação*, 4.ª ed., S. Paulo, Brasil: Ibrasa.

Tuckman, B. (1978). *Conducting Educational Research*, 2nd ed., N.Y., U.S.A.: Harcourt Brace Jovanovich, Inc.

4

TRATAMENTO DE RESULTADOS DE TESTES E QUESTIONÁRIOS



Apesar dos métodos da estatística descritiva se basearem em conceitos matemáticos simples, a própria experiência do autor, tem ajudado a verificar que alguns professores têm, por vezes, dificuldades nestes aspectos, quando pretendem realizar o trabalho indispensável de tratamento de resultados obtidos pelos seus alunos.

Dispensamo-nos de explicitar os procedimentos relativos ao cálculo de percentagens, pois nesta área, não é realmente crível, que algum professor necessite de informação básica.

Poderemos dizer que o tratamento de resultados faz-se, fundamentalmente, para determinar dois aspectos: a dispersão dos resultados e a tendência central dos mesmos.

Medidas de tendência central

Existem três tipos fundamentais de medidas de tendência central: moda, mediana e média.

Moda

A moda é o resultado que se repete com mais frequência. Vejamos um exemplo (Fig. 7).

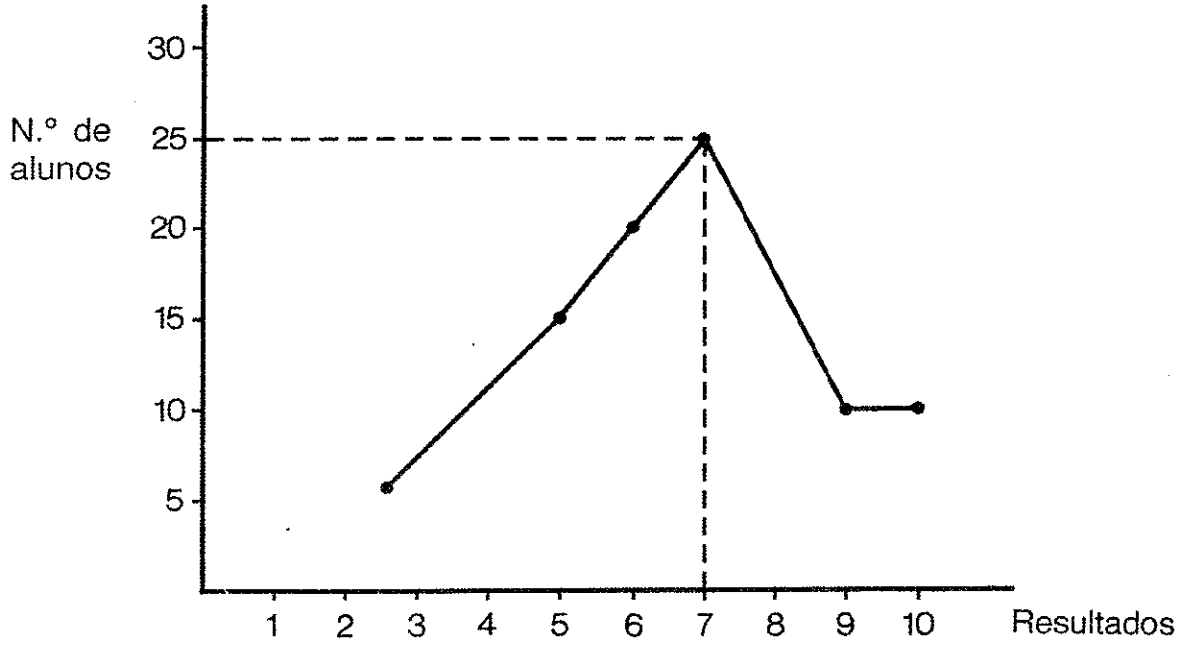


Fig. 7 — *Distribuição de resultados num teste*

Analisando o gráfico verificamos que o resultado que foi obtido por mais alunos foi 7. Esta é a moda.

Esta medida tem sido pouco utilizada por professores no tratamento dos resultados dos seus alunos. No entanto ela é importante e significativa, contrariamente ao que aparece em grande número de obras da especialidade.

Para professores que utilizam um sistema de pontuação diferenciado nos seus testes, ela na realidade tem pouca importância, sendo a média uma medida mais significativa. No entanto, quando o professor trabalha com objectivos definidos, aos quais se atribui igual importância, interessando portanto, saber, quantos objectivos foram atingidos, é a moda uma medida essencial. Permitindo-lhe a curva de distribuição de resultados, verificar a situação do total dos alunos, face ao total dos objectivos testados, a moda indica-lhe qual o grau de consecução dos objectivos obtidos por mais alunos, permitindo analisar o ponto de mestria alcançado.

Mediana

É aquilo que podemos designar pelo ponto médio da distribuição de resultados.

Assim 50% dos resultados estão acima desse ponto e 50% abaixo dele.

Suponhamos que temos uma turma de 21 alunos e que os resultados dos seus testes foram expressos numa escala de 0-100. Colocando os resultados numa tabela ascendente, teremos, por exemplo:

1	100
2	100
3	95
4	95
5	90
6	80
7	80
8	80
9	70
10	70
11	65
10	60
9	60
8	55
7	55
6	50
5	50
4	45
3	40
2	35
1	25

A mediana é portanto 65. Acima deste resultado existem 10 outros e abaixo também outros 10. A mediana é portanto o 11.º resultado (disposto em escala ascendente) que é o ponto médio.

Quando o número total de alunos é par (por exemplo 20 em vez de 21), a mediana será a nota média das notas centrais (a média entre o 10.º e o 11.º resultado).

A mediana não é uma medida muito usada, pois não tem grande utilidade prática para o professor, mas chama-se a atenção para as considerações feitas acerca da moda, pois, em parte também se aplicam a esta medida, além de ser uma medida auxiliar da média, em certos casos, como veremos a seguir.

Média aritmética

A média aritmética é uma medida que nos dá o rendimento médio de um grupo num teste. A média é calculada com a fórmula:

$$M = \frac{\Sigma X}{N}$$

sendo:

M – Média

ΣX – Soma dos resultados dos alunos

N – Número de alunos

A média é o indicador mais usado pelos professores, mas, no entanto, é insuficiente para extrair indicações necessárias dos dados. Vejamos por exemplo o caso em que existem valores extremos, muito afastados da média e que a distorcem, levando o professor a não tirar as ilações correctas. Um exemplo:

Resultados de alunos num teste:

30, 30, 35, 40, 40, 45, 45, 90, 95

Neste caso a média aritmética é 50. No entanto ela pouco significa porque, na realidade só há dois resultados acima da média, estando todos os outros abaixo. Os resultados acima estão, no entanto, nos extremos da tabela o que «inflaciona» a realidade do grupo. Neste caso, será correcto fazer o cálculo da mediana que é 40 e que espelha mais claramente a realidade do grupo.

Sobre a média convirá ainda dizer que ela será útil, nomeadamente quando se trabalha com pontuações diferenciadas nas perguntas e cada uma não corresponde a um objectivo, mas sim o teste a uma selecção da matéria. Quando o professor trabalha com objectivos divididos em níveis e faz corresponder a cada um, um item no teste, a média é um indicador de pouca utilidade, sendo mais esclarecedor o uso da moda e da mediana como atrás foi dito.

Medidas de dispersão

Ao professor não interessa só analisar os pontos médios e centrais dos resultados dos seus alunos, mas também a sua variabilidade, isto é, a sua distribuição ou dispersão pela escala utilizada.

Verificar em que medida os resultados variam uns dos outros, do mais baixo ao mais alto, tem grande interesse para a planificação e acção dos professores como veremos mais adiante.

Das medidas de dispersão a mais útil e importante para utilização dos professores na avaliação da aprendizagem é o desvio-padrão.

Desvio-padrão

O desvio-padrão é um indicador do grau de homogeneidade do grupo de resultados.

Para calcular o desvio-padrão utiliza-se o seguinte procedimento matemático:

$$DP = \sqrt{\frac{\Sigma X^2}{N}}$$

Sendo:

DP – desvio-padrão

X – diferença entre cada resultado e a média do grupo

N – número total de resultados

Vejamos um exemplo para concretizar:

Resultado de um grupo de 10 alunos num teste:

x (resultados)	X (diferença para a média)	X^2 (quadrado da diferença)
8	− 1,3	1,69
5	− 4,3	18,49
2	− 7,3	53,29
9	− 0,3	0,09
9	− 0,3	0,09
10	+ 0,7	0,49
15	+ 5,7	32,49
14	+ 4,7	22,09
10	+ 0,7	0,49
11	+ 1,7	2,89
Média = 9,3		Soma = 132,10

$$D.P. = \sqrt{\frac{132,10}{10}} = \sqrt{13,21} = \mathbf{3,63}$$

Este cálculo, apesar de mostrar muitos números, implica apenas alguns minutos ao professor, após listar os resultados obtidos. De notar, que muitas calculadoras manuais calculam o desvio padrão, a partir da listagem dos resultados, sem necessidade de cálculos adicionais.

A utilidade do desvio-padrão para o professor, distingue-se em dois aspectos: delimitação de intervalos (na óptica da curva normal), comparação de resultados dos alunos face aos intervalos e comparação dos resultados de dois grupos.

Os resultados situados entre + 1 D.P e − 1 D.P definem um intervalo que pode considerar-se normal relativamente ao grupo. Assim, no exemplo que demos atrás, os resultados situados entre 5,67 e 12,93 poderão considerar-se normais relativamente ao conjunto de resultados. Este cálculo permite pois, analisar o resultado de cada aluno face aos resultados globais da turma, bem como distinguir grupos dentro da turma (por exemplo o grupo mais fraco — com resultados abaixo do nível de − 1 D.P., o grupo normal — com resultados em − 1 D.P e + 1 D.P, e o grupo mais forte — com resultados acima de + 1 D.P).

Com estes dados o professor pode reajustar planos e metodologias de acordo com os grupos de alunos, nomeadamente no que respeita a actividades de remediação e enriquecimento. Por outro lado o cálculo da dispersão dá-lhe uma medida da homogeneidade da turma, aspecto que tanto afecta o decorrer do processo de ensino-aprendizagem. O professor pode verificar se cada uma das suas turmas é muito ou pouco heterogénea o que lhe permite definir a tempo objectivos e estratégias adequadas, face a essa realidade. Diremos mesmo, que o cálculo da dispersão, nomeadamente, dos resultados no início do ano, é um factor absolutamente fundamental para uma das funções de avaliação — a orientação.

Do mesmo modo que o desvio-padrão permite definir grupos dentro de uma turma, permite igualmente comparar os resultados de turmas diferentes, podendo o professor delinear de modo mais correcto e realista os seus objectivos e estratégias, pois é um dado essencial para a previsão dos resultados esperados e para a possível individualização do ensino.

BIBLIOGRAFIA:

Baquero, G. (1973). *Métodos de Pesquisa Pedagógica*, S. Paulo, Brasil: Edições Loyola.

Bartz, A.E. (1979). *Descriptive Statistics for Education and Behavorial Sciences*, 5th ed., Minneapolis, U.S.A.: Burgess Publ. Company.

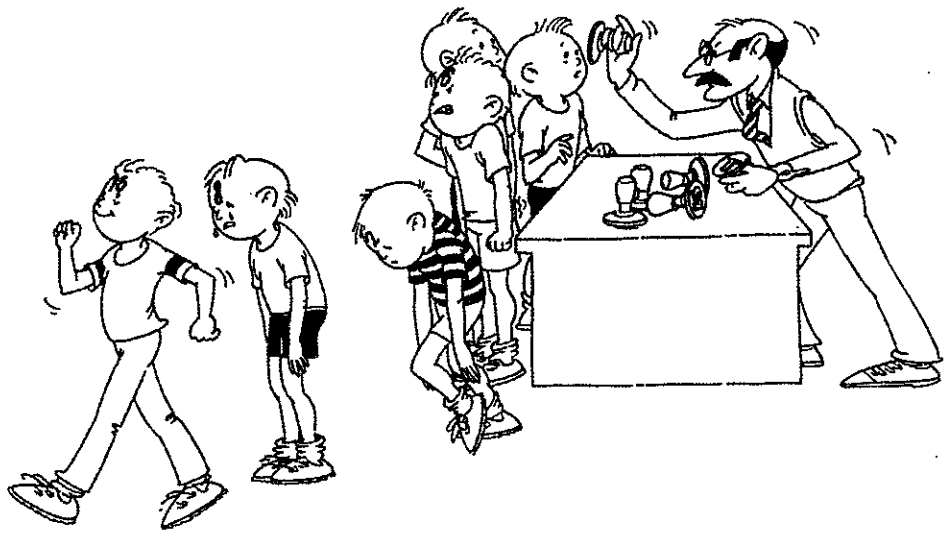
Franzblau, A. (1979). Uses of the Standard Deviation, in McFARLAND, S. e HEREFORD, C. (ed.), *Statistics and Measurement in the Classroom*, Dubuque, Iowa, U.S.A.: W.MN. C. Brown Company Publ.

Levin, J. (1985). *Estatística Aplicada às Ciências Humanas*, 2.ª ed., S. Paulo, Brasil: Editora Harper e Row do Brasil.

Nazareth, J.M. (1981). *Introdução aos Métodos Quantitativos em Ciências Sociais*, Lisboa: Universidade Nova de Lisboa.

5

A CORRECÇÃO E A PONTUAÇÃO DOS TESTES E A CLASSIFICAÇÃO DOS ALUNOS



No fim de um ano escolar, de um período lectivo, ou de uma acção específica de formação, o professor atribui uma classificação final aos respectivos alunos. Essa classificação, expressa numa escala quantitativa ou qualitativa tem como objectivos:

1 — Expressar de forma simplificada o nível de aprendizagem global atingido pelo aluno;
2 — Permitir estabelecer comparações entre níveis de aprendizagem dos diversos alunos.

A atribuição dessa classificação (que entre outros aspectos, vai determinar o sucesso/insucesso do aluno) é portanto o ponto culminante do processo de avaliação de aprendizagem e a sua importância é, como sabemos, decisiva. Por isso mesmo, requere-se nesse processo o maior rigor e objectividade. Requisitos que, como já vimos, são os mais importantes em todo o processo de avaliação, desde a construção dos instrumentos à sua aplicação, à correcção e pontuação dos resultados e ao tratamento dos mesmos.

Correcção e pontuação dos resultados

Um problema importante e que se coloca a muitos professores é a escala de pontuação a utilizar na cotação das perguntas de um teste e do respectivo nível de resposta. É evidente que a cotação das perguntas bem como a pontuação das respostas na escala escolhida está estritamente dependente de toda a planificação do ensino previamente elaborada e que implica actuações diferentes na avaliação por parte do professor, consoante a forma como planificou o ensino-aprendizagem. Assim, de acordo com a perspectiva de organização do ensino, deverão estar as técnicas de cotação e pontuação dos testes e restantes instrumentos de avaliação.

Cotação e escalas

Durante muitos anos vigorou em Portugal, uma escala uniformizada de classificação — a escala de 0 a 20 valores.
Com a evolução do sistema, assistiu-se ao aparecimento da escala de 1 a 5 nos ensinos preparatório e secundário (geral unificado), ao desaparecimento de escalas quantitativas no ensino primário, mantendo-se a escala 0 — 20 no ensino secundário (complementar) e superior (onde existiram durante um curto período de tempo, escalas qualitativas).

A partir de determinado momento, disseminou-se também a escala de 0 — 100 para pontuação de testes, na qual se consideram intervalos, aos quais se fazem corresponder os níveis de 1 a 5 (ou de 0 a 20), para classificação final do aluno.

Esta multiplicitude de forma, não acompanhada pela actualização respectiva dos professores, veio criar entropia no sistema escolar, onde se encontram agora professores a utilizar de forma indiferenciada e cruzada todas as formas possíveis. O que em si, não constitui qualquer erro, só o sendo quando essa utilização não é feita de forma adequada e consciente.

Sobre as escalas, haverá que dizer que qualquer resultado expresso numa escala de intervalo (1) é passível de ser convertido noutro, expresso em outra escala de intervalo. Portanto, quanto a esse aspecto, não decorrem problemas da utilização de escalas diferenciadas na pontuação dos resultados dos alunos, desde que o professor utilize a mesma escala, sempre com o mesmo grupo de alunos.

O problema coloca-se quando o professor quer atribuir a cada item de um teste ou outro instrumento, uma cotação, no contexto de todo o teste. Está em causa, neste caso a *importância relativa* de cada item. Como cada item diz respeito a um objectivo, é, portanto, a importância relativa de cada objectivo que é necessário estabelecer. O que nos conduz novamente à noção de pedagogia para a mestria.

Na perspectiva desta teoria pedagógica, todos os objectivos são igualmente importantes. Logo, todos têm igual cotação (além de não haver formas intermédias de pontuação da resposta, como veremos mais adiante). O nível da aprendizagem é dado pela quantidade e/ou qualidade dos objectivos alcançados.

Numa perspectiva pedagógica mais tradicional, as questões de um teste são hierarquizadas, por ordem de importância, atribuindo-se em consequência uma cotação que estabeleça essa hierarquia. Neste caso, o aspecto fundamental a ter em conta para para a cotação é: *a importância relativa do objectivo testado por cada item.* Aos de maior importância, corresponderá maior cotação. Faz-se notar, uma vez mais, que importância não quer dizer dificuldade. A importância de um objectivo não está forçosamente relacionada (nem directa, nem inversamente) com a dificuldade do item que testa a sua consecução. Além de que o índice de dificuldade real do item, só pode ser verificado após a sua aplicação, como se viu atrás, e o que o professor pode ter antes é somente a previsão da dificuldade.

Por vezes para o professor, é, no entanto, difícil definir a importância dos objectivos em causa. Para isto existem dois indicadores que podem ser utilizados:

1 — A ênfase (tempo e esforço) colocada nas actividades de ensino-aprendizagem dos diversos objectivos.
(A maior ênfase no ensino deve corresponder maior importância na avaliação).

(1) Existem três tipos fundamentais de escalas para a utilização em avaliação: nominais, ordinais e de intervalo. As escalas nominais são aquelas em que não existe qualquer razão de ordem matemática entre os vários níveis da escala. Por exemplo: Sim Não Talvez.
Escalas ordinais são as que exprimem uma relação de ordem entre os diversos níveis. Por exemplo: 1.º, 2.º, 3.º, 4.º, etc.
Escalas de intervalo são aquelas em que a distância entre cada dois níveis sequenciais é sempre constante: Por exemplo: 0-100%.

2 – A medida em que cada objectivo constitui um pré-requisito para outros que serão abordados em fase posterior.
(Os objectivos que constituem pré-requisitos de aprendizagem a fazer posteriormente, são mais importantes do que os que não possuem tal característica).

Critérios de sucesso e pontuação

Para além de cotação dos itens, outro aspecto é importante de analisar: a pontuação da resposta dada pelo aluno. Uma vez mais voltamos a ter em causa dois procedimentos: o que respeita à pedagogia para a mestria e à avaliação criterial e o que respeita à pedagogia mais tradicional.

Um dos pressupostos fundamentais da primeira, é o de que a cada objectivo (e respectivos itens de avaliação) corresponde um critério de sucesso previamente estabelecido. Se o aluno cumpre o critério de sucesso, considera-se atingido o respectivo objectivo, se não cumpre o critério considera-se o objectivo não atingido. Não há, portanto, graus intermédios de resposta, aos quais corresponde uma pontuação intermédia. Não tem portanto interesse a pontuação de perguntas e respostas, porque de certo modo, ela nem existe.

O que determina o nível de aprendizagem do aluno é a quantidade ou o tipo de objectivos atingidos (consoante se definem os níveis por quantidades de objectivos atingidos ou pela identificação de todos os objectivos que estão incluidos em cada nível).

Neste tipo de prática, é normal fazer uma separação qualitativa entre o nível mínimo e o nível de desenvolvimento, isto é, seleccionar os objectivos de nível mínimo e separá-los dos de desenvolvimento, e dentro destes fazer uma abordagem quantitativa para os diversos níveis de sucesso.

Na prática mais tradicional, não se considera a noção de «mestria» (em que o indivíduo atinge ou não atinge o objectivo) aceitando-se portanto, que a resposta do aluno pode variar num intervalo de correcção. Não está forçosamente correcta ou incorrecta de modo absoluto, podendo revelar situações intermédias. Não existe, portanto, um critério absoluto de sucesso, mas sim uma escala de correcção e adequação da resposta dada.

Correcção e pontuação de testes objectivos

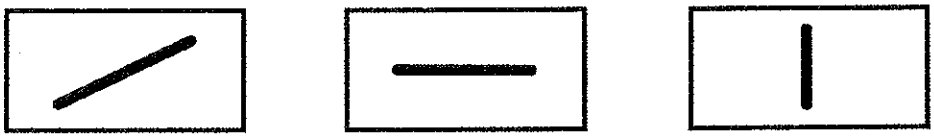
Na perspectiva da mestria, os testes objectivos são os únicos que devem ser usados, pois que não introduzem problemas para a definição do sucesso nas respostas. Quando a resposta é composta por várias partes (casos de resposta curta — identificação, completamento de frases, verdadeiro — falso, combinação, etc.) haverá que considerar previamente quais ou quantas as partes que deverão ser obrigato-

riamente respondidas correctamente para se considerar a resposta satisfatória, isto é, considerar que o aluno atingiu o objectivo.

Fora da perspectiva da pedagogia para a mestria atribui-se uma cotação à questão. Neste casos essa cotação deve ser previamente repartida pelas diversas partes da resposta, tendo em conta, uma vez mais, o critério da importância de cada parte no contexto da resposta. Às partes mais importantes deverá corresponder maior cotação, e às menos importantes, menor cotação. Se todas as partes são igualmente importantes, logicamente a cotação deverá ser igualmente repartida.

Vejamos um exemplo:

Questão: Identifica as linhas apresentadas, quanto à sua orientação no espaço:



Designação: _____ _____ _____

Suponhamos que a cotação do item era de 15 pontos. Se considerarmos que qualquer das três partes tem igual importância, a cada identificação correcta corresponderão 5 pontos. Mas se considerarmos a hipótese que, no contexto do ensino, no momento da realização do teste, a correcta identificação de linhas oblíquas era mais importante relativamente a outras orientações, far-se-ia corresponder à primeira a cotação de 7 pontos e 4 a cada uma das outras.

Tratando-se de itens cuja resposta não é composta por diversas partes, sendo, portanto, única e exclusiva, não há qualquer dúvida quer em considerar o objectivo atingido ou não, quer em atribuir toda a pontuação ou nenhuma.

Exemplo: O raciocínio probabilístico é característico de um dos estádios de desenvolvimento cognitivo descrito por Piaget. Identifique-o na lista seguinte:

1. _____ Operacional concreto
2. _____ Operacional analítico
3. _____ Operacional formal
4. _____ Proporcional

Só uma resposta é correcta, pelo que qualquer que seja a pontuação ela será atribuída por inteiro aos alunos que seleccionarem a hipótese correcta, não sendo atribuída qualquer pontuação aos que seleccionarem qualquer outra.

Correcção e pontuação de testes não-objectivos

Como já se disse, na pedagogia para a mestria, evita-se o uso de testes não objectivos. Quando se torna indispensável a sua utilização, deve-se, de igual modo,

definir previamente o critério de sucesso. Tal torna-se mais fácil quando se quantifica a pergunta (por ex: indicar número de exemplos, de argumentos, de justificações, etc, a apresentar na resposta).

Quando não estão em jogo quantidades na resposta, mas sim conceitos, princípios, factos, etc., deve previamente o professor fixar os termos, conceitos, demonstrações, etc. que a resposta deve conter para ser considerado atingido o objectivo.

Quando se trabalha com cotações, existem três procedimentos complementares que podem ser utilizados na correcção e pontuação de testes não-objectivos.

Qualidade absoluta da resposta — O professor deverá elaborar uma resposta modelo ao item em causa. Após essa elaboração pontuará três aspectos da resposta: (1) o aspecto quantitativo, isto é, o número de dados apresentados (exemplos, razões, argumentos, etc); (2) o aspecto qualitativo, isto é, a correcção e/ou a coerência dos dados apresentados, e (3) a estrutura global da resposta, isto é, a ligação entre as diversas partes, a sua coerência e a sua clareza face à questão respectiva.

Para melhor esclarecimento, vejamos um exemplo:

Questão: Estabelece uma relação entre a estrutura dos componentes do sistema circulatório e a respectiva função no transporte dos princípios nutritivos (cotação: 30 pontos).

Resposta-modelo — Componentes:

	Órgãos	Estrutura	Função
Vasos Sanguíneos	Coração	Muscular, fibrosa	Bombeamento
	Artérias	Parede elástica com válvulas	Progressão da corrente sanguínea; evitar o retorno (refluxo)
	Veias	Parede não elástica	Passagem de retorno
	Capilares	Dimensão reduzida e parede fina	Distribuição do sangue por todos os locais
Sangue	Plasma	Fluido	Recepção de nutrientes e circulação pelos vasos e espaços intercelulares
	Glóbulos vermelhos	Células com hemoglobina	Recepção do oxigénio e dióxido de carbono
	Plaquetas	Fragmentos celulares	Coagulação no rompimento de vasos
	Glóbulos brancos	Células de forma variável	Ataque a microrganismos que penetram na corrente sanguínea

Consideremos que para o nível de escolaridade em causa, este era um esquema possível de identificação das componentes da resposta. Analisemos agora uma hipótese de pontuação (para um total de 30 pontos):

Identificação de cada componente — 1 ponto (total 8 pontos)
Identificação de cada relação estrutura-função — 2 pontos (total 16 pontos)
Estrutura global da resposta:
Considerar todas as componentes e relações: 2 pontos
Organização da resposta de forma agrupada: 2 pontos
Clareza na definição das relações estrurura-função: 2 pontos (total 6 pontos)

Qualidade relativa da resposta — Diz respeito à comparação das respostas de todos os alunos. Nesta perspectiva o professor analisa primeiramente as respostas de todos os alunos. Seguidamente escolhe a que considera melhor, atribuindo a essa o total da pontuação (ou uma parte substancial desse total, se considerar que mesmo essa resposta não preenche em absoluto todos os requisitos fixados). Após isto o professor escolhe a pior resposta (não considerando as omissões) e atribui a esta a pontuação mínima. Seguidamente o professor hierarquiza todas as respostas entre as duas citadas e distribui níveis de pontuação ao longo dessa hierarquia.

Correcção múltipla — Este é um procedimento de que já se falou atrás, a propósito da validade de conteúdo dos testes. Trata-se de submeter as respostas à apreciação de dois ou mais professores. É um procedimento que tem alguns inconvenientes práticos, mas que se torna perfeitamente possível quando exista um sistema de trabalho em equipa numa escola. Para além disso, tal sistema é, no entanto indispensável quando se trate de analisar testes não-objectivos como instrumento de avaliação geral em toda a escola. Em tal caso, deve-se sempre adoptar este procedimento.

A adopção da correcção múltipla não obsta, no entanto, aos procedimentos já descritos tendo em vista as qualidades absolutas e relativas das respostas. Como dissemos, no início, estes três procedimentos são complementares, e de modo nenhum mutuamente exclusivos. Em termos de prática geral, diremos que os relativos à cotação com vista à qualidade absoluta da resposta, são indispensáveis em qualquer caso, podendo ser complementados por qualquer dos outros dois, ou por ambos.

Correcção e pontuação de registos de observação

Os registos de observação, como já se disse, podem ser de dois tipos: listas de verificação e grelhas de análise.

As listas de verificação respeitam à observação da realização de acções. Respeitam, portanto ao que, por vezes, também se designa por testes de desempenho. A sua utilização faz-se principalmente em áreas como a educação física e a educação visual e manual, onde frequentemente, se pretende verificar realizações práticas dos

alunos, Também, por vezes, em áreas de actividade prática das ciências da natureza (biologia, física, química, etc.) a sua utilização é necessária.

No respeitante à cotação dos diversos itens duma lista de verificação, haverá também que considerar duas perspectivas. Na respeitante à pedagogia para a mestria o que interessa é o produto de aprendizagem, isto é, o comportamento terminal, ou em casos mais complexos a realização bem sucedida ou não da acção a que cada item da lista diz respeito.

Tratando-se dum sistema de cotação haverá que ter em conta dois aspectos: (1) O que se relaciona com a cotação relativa de cada item, tendo em conta o total da lista, e (2) o que respeita à pontuação do nível de realização de cada acção.

Vejamos um exemplo:

Questão: Usando a folha de cartolina que te foi fornecida, constrói um cubo, cuja aresta meça 5 cm.

Lista de verificação a utilizar pelo professor:

(9) (0). Desenha um quadrado com 5 cm de lado, segundo as regras de construção estudadas.

(9) (0). Desenha os restantes quadrados correspondentes às diversas faces do cubo, utilizando o primeiro como referência central e usando as mesmas regras de construção.

(9). Desenha os espaços para colagem das faces.

(10) (0). Utiliza a régua e o esquadro de forma correcta.

(9) (0). É rigoroso nas medições efectuadas.

(9). Recorta o desenho obtido.

(9) (0). É rigoroso na execução do recorte.

(9). Utiliza o x-acto para fazer um corte de pequena profundidade nas arestas de dobragem para facilitar esta.

(9). Executa a dobragem da cartolina recortada.

(9). Faz a colagem final, de modo a obter o cubo.

(9) (0). Executa a colagem de forma rigorosa, obtendo faces perfeitamente quadradas e arestas bem definidas.

(Entre parêntesis indica-se uma possível cotação)

Neste exemplo hipoteticamente cotado para um total de 100 pontos, estão identificados onze itens de avaliação. No exemplo estão todos cotados de igual modo — 9 pontos — à excepção de um que tem 10. Claro que o professor pode cotá-los de

forma diferenciada, em função da sua importância, como já disse para os testes. Neste caso considera-se hipoteticamente, que são todos igualmente importantes.

Como se verifica estão considerados dois tipos de itens. Os que se referem a uma realização do aluno, que não tem formas intermédias, e, portanto, ele obtém o total da respectiva pontuação ou nenhum, e os que se referem a uma realização cuja qualidade pode variar numa escala, a que corresponde a atribuição entre o máximo e o mínimo.

Além das listas de verificação, os registos de observação podem também tomar a forma de grelhas de análise. Destinam-se estas a registar as frequências de ocorrências de determinados comportamentos dos alunos. Em alguns pretende-se que o aluno atinja uma frequência elevada, como em (no exemplo que foi dado atrás, quando se descreveram as características de uma grelha de análise): «Arruma o material» e noutros pretende-se a ausência de manifestação como nível ideal, como por exemplo em: «Interrompe os colegas». Para estas frequências devem ser fixados níveis correspondentes numa escala de pontuação, que, neste caso, é muito aconselhável que possua poucos intervalos (ex: 1 a 5) por razões de aplicabilidade prática. Assim, por exemplo, um aluno que arruma sempre o material teria 5 neste item e o que nunca arruma teria 1, fixando frequências intermédias para os níveis 2, 3, 4, sendo o 2 para o que raramente (só uma vez ou duas) arrumou o material e o 3 e o 4 para os que costumam arrumar o material não o fazendo, no entanto, por vezes.

Classificação dos alunos

No fim do período ou ano, haverá que transformar todos estes resultados, num único, expresso no nível de uma escala numérica ou qualitativa. É o trabalho de tentar resumir toda a avaliação de aprendizagem num número.

Haverá que analisar e pesar, portanto, todos os resultados obtidos de forma a conjugá-los num só resultado.

Há algumas regras absolutamente indispensáveis, a ter em conta, neste momento:

1 – Só os dados objectivos, ou passíveis de serem objectivados, devem ser usados para o cálculo da classificação.

2 – Tudo o que foi avaliado e vai ser tido em conta para este efeito, deverá ter sido do prévio conhecimento dos alunos. O professor não pode usar dados de avaliação de aspectos de que não deu prévio conhecimento aos alunos, porque estará a cometer uma grosseira desonestidade. O aluno tem que saber claramente o que o professor espera do seu trabalho e da sua aprendizagem.

3 – Só pode ser avaliado (e considerado aqui) aquilo que foi objecto de ensino. Isto é, o professor não pode avaliar e usar resultados da avaliação de aspectos que não fizeram parte do seu ensino. Esta questão coloca-se frequentemente no que respeita a comportamento socio-afectivos, aquilo que gene-

ricamente muitos professores designam pelo «comportamento» do aluno. Por vezes os professores não informam os alunos sobre que tipo de «comportamento» esperam dele, nem fazem desses comportamentos objectivo de ensino, no entanto, servem-se de resultados a maioria das vezes não recolhidos de forma objectiva e sistemática, para alterarem o nível de resultados final da aprendizagem dos alunos obtido através dos instrumentos de avaliação utilizados. Por exemplo, se o professor considera como resultado importante da aprendizagem que os alunos não devem interromper o professor, sem pedirem previamente a palavra, deve informar os seus alunos disso, deve dedicar tempo de ensino à aprendizagem e disseminação desse comportamento e deve avaliá-lo da forma mais objectiva que lhe seja possível. Deste modo poderá considerá-lo uma componente no tratamento dos resultados da avaliação, com vista à atribuição de uma classificação.

Utilização da escala de 1 a 5

Nesta escala os níveis 1 e 2 significam insucesso e os restantes representam o sucesso. A primeira questão coloca-se com a utilização do nível 1. Sendo aritmeticamente possível, não nos parece, no entanto, aconselhável. A discriminação em níveis negativos dos resultados da aprendizagem, corresponde à utilização do «reforço negativo» e do «castigo» da teoria de comportamento condicionado de Skimer. E se a utilização do «reforço positivo» é feita por quase todos os professores e alguns estudos mostram bons resultados da sua utilização, o mesmo não se passa com o «castigo». Nem as investigações existentes o demonstram, nem é de querer que a discriminação negativa possa produzir boas consequências pedagógicas, sendo mais provável obter humilhações que irão provocar o desenvolvimento de atitudes negativas face à aprendizagem, ao professor e à escola. É uma forma errada de os professores desafiarem o «ego» dos alunos.

Assim restam-nos o 2, 3, 4 e 5.

Na perspectiva que desenvolvemos sobre a definição dos níveis de aprendizagem é fácil a atribuição do respectivo nível aos alunos. Assim os alunos que não atingiram os objectivos considerados no nível mínimo situar-se-ão no 2. Dos restantes, aqueles que atingiram o nível mínimo, mas atingiram menos de 25 a 30% dos objectivos do nível de desenvolvimento, situar-se-ão no 3. Os que atingiram 30 a 60-65% desses objectivos situar-se-ão no 4 e os restantes no 5. Simples, fácil e rigoroso.

Trabalhando com pontuações a situação é um pouco mais complexa. Suponhamos que o professor possui 4 resultados de avaliação respeitante a um aluno. Consideremos os resultados expressos na escala de 0 a 100.

1.° Teste — 64

Trabalho prático (avaliado através de lista de verificação) — 80

Grelha de análise de comportamento — 50

2.° Teste — 48

Primeiramente haverá que atribuir pesos a cada um destes resultados. Em princípio, se os testes foram válidos, isto é, se mediram realmente os resultados de aprendizagem efectuada, deverão normalmente considerar-se com peso mais elevado. Não quer isto dizer, que face às situações reais de ensino e aprendizagem a decisão não possa ser outra. Suponhamos que se atribuia peso 2 ao 2.° teste e peso 1 a cada um dos outros resultados. Teríamos então:

$$\frac{64 + 80 + 50 + 48 \times 2}{5} = 58$$

Esta seria portanto a média ponderada de resultados obtidos. Resta agora enquadrá-la na escala de 1 a 5. Segundo indicações do Ministério da Educação em provas de exame, deve considerar-se a seguinte correspondência de intervalos 1 — 0 a 29; 2 — 30 a 49; 3 — 50 a 74; 4 — 75 a 90; 5 — 91 a 100. Será uma correspondência tão boa, como outra qualquer. Esta correspondência ou qualquer outra não obsta ao aparecimento dos chamados casos — fronteira. Isto é, dos alunos cuja média de resultados se encontra no limite de um nível. Em tais casos poderão considerar-se como indicadores, os resultados obtidos nos instrumentos de avaliação de peso maior, o ritmo de progressão do aluno na aprendizagem, ao longo do tempo, o cumprimento dos prazos fixados para apresentação de trabalhos ou outras realizações, etc. Aqui, à falta de melhor, o professor poderá considerar factores subjectivos, tentando objectivá-los então. Se ainda assim, as dúvidas se mantiverem, o professor não deverá prejudicar o aluno e fará mais uma utilização do «reforço positivo».

A existência destes casos é frequente e não constitui situação anómala. mas a sua persistência sim. Isto porque ao acontecer, o professor deverá usar todas as técnicas possíveis, quer por um lado para fazer progredir o aluno na aprendizagem e «puxá-lo» para o nível seguinte, quer por outro utilizando, se necessário, mais instrumentos e mais rigorosos, para medir a aprendizagem efectuada.

A existência de casos-fronteira, em número elevado no final do ano, constitui para um professor indicador seguro da baixa qualidade dos seus procedimentos e instrumentos de avaliação.

Utilização da escala 0 a 20

Os procedimentos a adoptar para a utilização desta escala, não diferem substancialmente dos já descritos. Relativamente ao nível mínimo e nível de desenvolvimento, através da consecução dos respectivos objectivos sabe-se que a fronteira para o nível mínimo será 10. Os onze níveis existentes entre 10 e 20 far-se-ão corresponder à percentagem de objectivos do nível de desenvolvimento atingidos. Se o professor considerar desejável e possível distinguir onze níveis diferentes de desenvolvimento, poderá fazê-lo, se o não considerar poderá fixar os níveis desejados e fazer corresponder cada um a um determinado valor na escala.

Tendo os resultados expressos em pontuações, haverá que determinar o peso de cada resultado e obter a média ponderada dos resultados. Após isso a transferência para a escala de 0 a 20 pode fazer-se simplesmente através da aplicação da razão entre as duas escalas. Assim, para o exemplo dado atrás (na escala de 0 a 100) bastará dividir o resultado da média ponderada por 5 e fazer a respectiva aproximação, se for caso disso. Assim seria 58 : 5 = 11,6 (*12* aprox.).

Trabalhando em pontuações com a escala de 0 a 100, que é o procedimento mais usual, bastará, portanto, fazer a respectiva divisão por 5 da média ponderada dos diversos resultados de avaliação, para fazer a respectiva redução à escala 0 a 20.

Pelo que foi dito ressalta, naturalmente, que a conversão de resultados expressos em escalas com grande número de intervalos para escalas com menor número de intervalos é sempre possível (e desejável).

A situação inversa é que não é desejável, nem possível muitas vezes, pelo menos de forma correcta. Isto é, por exemplo, um resultado expresso nas escalas de 0 a 5 não é convertível na escala de 0 a 20. Para além das considerações relativas à (in)correcção matemática de tal procedimento, haverá que considerar que a escala de 1 a 5 não é verdadeiramente uma escala de intervalo, mas mais uma escala nominal. Quer isto dizer que entre os diversos graus da escala (entre 1 e 2 entre 2 e 3, etc.) não existem intervalos exactamente iguais, o que não acontece, em princípio, com as escalas de 0 a 20 e de 0 a 100.

BIBLIOGRAFIA:

Bonboir. A. (1976). *Como Avaliar os Alunos,* Lisboa: Seara Nova.

Berck, R. A. (1986). A consumer's Guide to Setting Performence Standart on Criterion--Referenced Tests.*Review of Educational Research, 56* 137-172

Goring, P. (1981). *Manual de Medições e Avaliação do Rendimento Escolar,* Coimbra: Livraria Almedina.

Hambleton, R. K. et. al. (1978). Criterion-Refereenced Testing and Measurement: a Review of Technical Issues and Developments. *Review of Educational Research, 48,* 1-47

Nikto, A. J. (1983), *Educational Tests and Measurement: An Introduction*, N.Y., U.S.A.: Harcourt Brace Jovanovich, Inc.

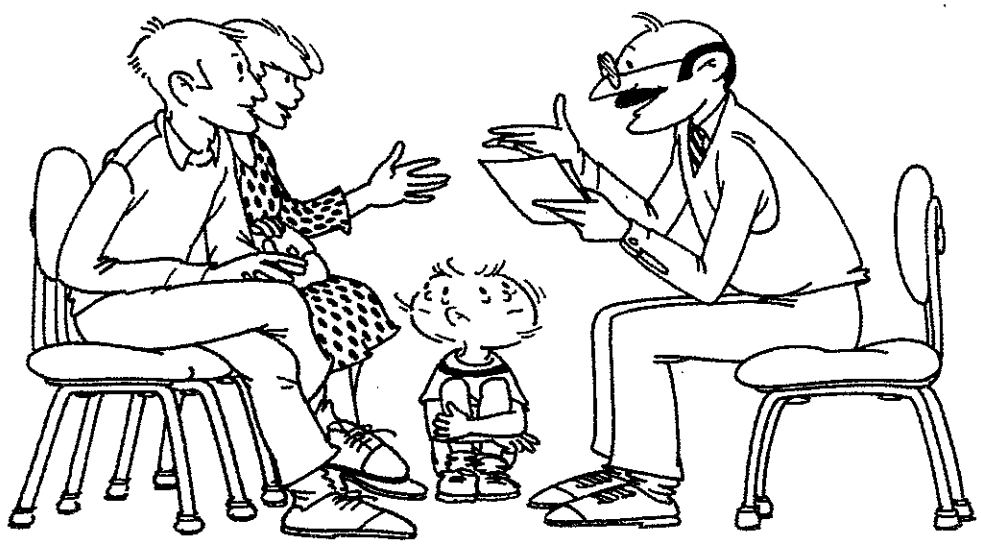
Noizet, G. e Caverni, J. P. (1985). *Psicologia da Avaliação Escolar,* Coimbra: Coimbra Editora.

Pidgeon, D. e Yates, A. (1976). *Evaluation y Medida del Rendimento Escolar,* Salamanca, Espanha: Ediciones Anaya

Tuckman, B. (1978). *Conducing Educational Research,* 2nd ed., N. Y., U.S.A.: Harcourt Brace Jovanovich, Inc.

6

A UTILIZAÇÃO DOS DADOS DE AVALIAÇÃO DA APRENDIZAGEM



O que ficou dito nos capítulos anteriores não pretende ser um texto exaustivo sobre avaliação, mas tão somente uma abordagem dos principais aspectos da avaliação da aprendizagem, numa perspectiva o mais prática possível, de modo a fornecer informações instrumentais que facilitem e, se possível, melhorem o trabalho diário do professor.

A avaliação é uma componente demasiado importante, para que possa ser deixada ao acaso, aos humores dos professores e alunos, ou, mesmo, exclusivamente ao «senso comum» e a esquemas de aperfeiçoamento baseados somente na tentativa e erro.

A sua importância deriva, não só, da necessidade do professor conhecer, com rigor, qual foi a aprendizagem real dos seus alunos, mas, também, e talvez principalmente, da utilização subsquente desse conhecimento. A utilização dos dados de avaliação da aprendizagem é um dos aspectos fulcrais de todo o sistema de ensino-aprendizagem. Ela define o nível de sucesso deste bem como os seus possíveis aperfeiçoamentos. Constitui também o melhor indicador de capacidade profissional do professor. Os dados de avaliação da aprendizagem são úteis na orientação e aconselhamento dos alunos, na selecção e modificação de metodologias, na avaliação dos programas, na previsão de resultados futuros e na investigação.

Orientação e aconselhamento

Nas nossas escolas não existe definida de modo claro a função de orientação e aconselhamento dos alunos e pais. De qualquer modo o professor deve ser sempre um orientador, e ao director de turma estão particularmente atribuídas funções nesse âmbito.

Os resultados da avaliação da aprendizagem, são, sem dúvida, o indicador mais óbvio da situação escolar do aluno e das acções a desenvolver por este e pelos pais no sentido de a melhorar, de seleccionar adequadamente opções de estudo, desenvolver campos de interesse, de obviar a insucessos previsíveis, em tempo útil, etc.

Caso típico é o relativo às informações aos pais, após o primeiro momento definido de avaliação. O mais importante não é informar os pais dos resultados objectivos dos seus filhos, mas sim, do que poderá ser feito por eles no sentido de esses resultados poderem vir a melhorar ao longo do ano. Nomeadamente no que respeita aos alunos que demonstrem especiais dificuldades nesta ou naquela disciplina. Sugestões sobre tempos de estudo, métodos de trabalhos, realização de tarefas em casa, devem ser dadas de forma específica, não só aos alunos, mas também aos respectivos pais.

A queixa de pouco interesse dos pais pelos resultados escolares dos filhos é muito ouvida nas nossas escolas. Mas, a verdade, é que a responsabilização dos pais pelos professores, também não é suficientemente acentuada. Sendo claro que não podem obrigar-se compulsivamente os pais a ir à escola, podem, no entanto, fazer-se chegar a estes, informações detalhadas sobre a aprendizagem dos filhos.

A prestação continuada de informações torna-se difícil para os professores, sendo feita em relação a todos os seus alunos. Mas, também não é necessária. Devem seleccionar-se os alunos com mais dificuldades, e para esses, ter esse tipo de atitude.

Durante algum tempo gerou-se a opinião que os professores não deveriam solicitar aos alunos a realização de trabalhos em casa, nem manter ligação de informação aos pais, através dos alunos. Nada de mais errado. Sobre os trabalhos em casa, falaremos a seguir e sobre as informações aos pais, haverá que perguntar qual o melhor e mais natural meio de ligação entre a escola e a família, do que o próprio aluno, que se inclui nos dois grupos?

As mensagens periódicas aos pais sobre o trabalho dos filhos, responsabilizam ambos na educação destes. Não deverá, nunca considerar-se que o director de turma é o único interlocutor com as famílias. O director de turma é o interlocutor privilegiado da relação da escola no seu todo, com a família, mas o professor é o interlocutor adequado no respeitante aos seus alunos, na sua área ou disciplina.

As informações e conselhos aos pais podem ser prestados de várias formas, mas a mais natural é através de mensagens orais ou escritas dadas aos alunos para transmissão aos pais. O professor pode, e deve, sempre que o entenda conveniente, escrever informações nos cadernos e dossiers do aluno, ou em qualquer outro local e solicitar a confirmação de conhecimento pelos pais. Não é uma prática antiquada ou punitiva, mas sim uma prática que aumenta a circulação da informação e a responsabilização das duas partes mais importantes na educação dos alunos: os pais e a escola.

Recuperação e enriquecimento

Alguns trabalhos têm provado a estreita correlação entre o primeiro e o último resultado dos alunos, no mesmo ano escolar. Isto é, a maioria dos alunos que têm classificação insuficiente ou negativa no início do ano, numa dada área ou disciplina, vêm a ter insucesso na mesma, no final do ano.

A confirmar-se esta pesquisa levanta-se um problema grave na escola. Os alunos não têm capacidade de recuperar na aprendizagem? O professor «marca» o insucesso dos alunos no início do ano? Ou o professor não cria condições e não dá oportunidades à recuperação dos alunos?

Como é lógico, cremos que a última hipótese é a mais verosímil. Nem todos os alunos obtêm bons resultados de aprendizagem no início do ano. No entanto, muitas vezes o professor continua calma e paulatinamente, «dando o programa», sem ter em atenção tal facto, provocando naturalmente o cada vez maior «afastamento para baixo» daqueles alunos.

Se é verdade que o objectivo do professor é levar cada um dos seus alunos a aprender o máximo possível, é igualmente verdade que, não trabalhando num sistema de ensino individualizado, a sua primeira obrigação é tentar levar todos os alunos a aprender o mínimo indispensável.

Coloca-se aqui, portanto, o problema da recuperação ou da remediação. O professor deve realizar acções de recuperação dos alunos com piores resultados, assim que obtém dados de avaliação da aprendizagem. Esta é uma das funções mais importantes da avaliação: o feed-back relativamente à planificação do ensino.

Muitas vezes, os professores queixam-se que não têm tempo para acções de recuperação, porque têm que continuar a «dar o programa». Mas «dar o programa» a quem? A progressão à toa pelo programa sem a recuperação de atrasos dos alunos, provoca atrasos cada vez maiores e progressivamente irrecuperáveis. Além de que a alegada falta de tempo, se deve, quase sempre, a planeamento deficiente, quando não inexistente. No seu planeamento o professor deve prever espaços para actividades de recuperação, porque sabe que inevitavelmente elas terão lugar. Esses espaços são simultaneamente utilizados em actividades de desenvolvimento ou enriquecimento, dos alunos com melhores resultados, o que obsta também à observação por vezes feita, de «durante esse tempo, o que faço aos que não necessitam de recuperar?». Uma aula, pode perfeitamente servir os dois desígnios. Enquanto o professor trabalha com o grupo de alunos em recuperação, distribui aos alunos do outro grupo tarefas e actividades de desenvolvimento ou enriquecimento.

J. Mateus e colaboradores (1984, 1985) desenvolveram uma metodologia de planificação e execução do ensino, cuja aplicação facilita imenso o trabalho do professor, no que respeita a este problema, além de sistematizar outros aspectos do desenvolvimento curricular. Esperamos que tais trabalhos venham brevemente a ser publicados, pois viriam facilitar o trabalho de muitos professores preocupados com a questão da remediação e enriquecimento dos seus alunos.

O trabalho para casa é outra estratégia muito ligada à recuperação e enriquecimento dos alunos. É aliás nesta perspectiva que ele pode ter uma boa utilidade. O professor não precisa de recomendar a realização de trabalhos em casa a todos os alunos, nem o mesmo tipo de trabalhos para todos. O professor deve escolher trabalhos adequados às necessidades de cada aluno.

A utilização desta metodologia dá a oportunidade ao aluno de recuperar ou de se enriquecer, mas de forma individualizada e adaptada às suas necessidades e desejos.

A própria vontade de realização dos trabalhos, constitui um bom indicador para o professor, para além da qualidade resultante dos mesmos.

Modificação dos métodos de ensino

No início deste capítulo referimos que os resultados da aprendizagem eram um bom indicador da capacidade profissional do professor. Não quer isto dizer que um professor cujos alunos obtiveram todos muito boas classificações seja forçosamente um «bom professor» e um outro cujos alunos obtiveram más classificações seja, inequivocamente, um «mau professor». No entanto, existem alguns indicadores que devem ser tidos em conta na avaliação dos resultados obtidos por um professor, no

que respeita às classificações dos seus alunos. Um indicador é o índice médio de insucesso. Este índice pode referir-se a dois níveis: o nacional e o de escola. Quando os resultados dos alunos, ultrapassarem tal índice, isto deverá significar o acender de uma luz de alerta para o professor.

Outro indicador refere-se ao nível médio da turma nas diversas disciplinas. A tal «luz de alerta» deve acender-se quando os resultados negativos numa disciplina ultrapassem claramente os obtidos nas outras disciplinas ou áreas.

O que fazer, portanto, em tais situações? Ou seja, como utilizar os dados da avaliação para melhorar a qualidade do ensino e da aprendizagem?

Comecemos pelo que não deve ser feito. O professor não deve, nunca, modificar classificações atribuídas, só para se situar em níveis aceitáveis. Isto além de uma questão técnica, é também uma questão deontológica. As classificações finais devem reflectir o nível de aprendizagem atingido pelo aluno, e como tal, o professor não deve modificar nada, só para mostrar um menor ou maior «número de negativas». A ocorrerem modificações elas só poderão resultar de outros factores, como adiante veremos. É importante dizer isto, porque, por vezes, nomeadamente alguns professores menos experientes ou menos preparados, são «induzidos» a alterar classificações, nas reuniões de avaliação, somente pela razão de não quererem mostrar resultados muito fora dos intervalos obtidos pelos restantes colegas. Isto deve-se a um aspecto de insegurança profissional, que é possível ser ultrapassado.

A mesma questão de segurança está em causa quando os professores tomam outro tipo de atitudes que também devem ser evitadas. Trata-se do professor cujos alunos obtiveram altos índices de resultados negativos e que «passeia por toda a escola o seu orgulho em produzir calamidades pedagógicas» e do professor que entra em desespero por verificar os maus resultados dos seus alunos, culpando-se irremediavelmente da sua incapacidade.

É evidente que qualquer das duas atitudes é muito pouco profissional e revela a tal falta de segurança. No primeiro por não ser capaz de avaliar o que faz e extrair as ilações necessárias e no segundo por não ser capaz de equacionar as razões dos resultados e planear a sua melhoria.

O verdadeiro problema não está nos resultados em si mesmos, mas na forma como foram atingidos e na actuação seguinte do professor face aos mesmos.

Quando os resultados obtidos não se enquadram nos níveis esperados e desejáveis o professor deve:

1 – Reanalisar os instrumentos de avaliação utilizados e tentar verificar se os mesmos estão verdadeiramente adequados, procurar erros ou omissões, rever todo o sistema de avaliação, de modo a certificar-se da existência ou inexistência de falhas;

2 – Caso não se verifiquem falhas nem nos instrumentos nem no sistema de avaliação, fazer a revisão de como decorreram as actividades de ensino-aprendizagem, comparando cada uma com o resultado obtido pelos alunos na avaliação dos objectivos que lhe respeitavam;

3 – Reanalisar os objectivos fixados, tendo em conta os resultados anteriores dos alunos, a organização desses objectivos, e sua sequência e as capacidades e conhecimentos que neles estão incluídas.

Em cada uma destas fases, quando o professor encontrar erros ou falhas deve planear as acções necessárias para as colmatar. Por exemplo:

— Voltar a testar os objectivos em causa, quando verificar erros no respectivo instrumento de avaliação;
— Voltar a realizar essas actividades de aprendizagem, quando se verifica uma boa avaliação, mas um erro prévio de ensino;
— Incluir na planificação da unidade seguinte, objectivos da anterior que estavam mal enquadrados, ou cujas actividades de ensino-aprendizagem foram mal conduzidas;
— Suprimir objectivos quando se tenha verificado a sua inutilidade em termos de aprendizagem.

Outras acções podem ser levadas a cabo, dependendo da situação concreta verificada. No entanto há duas regras, que é indipensável respeitar:

1 – A avaliação do ensino deve ser constante, de modo a que os eventuais erros cometidos, possam ir sendo remediados ao longo do tempo, assegurando um caminho de aprendizagem, progressivamente mais seguro. Isto permitirá que o professor possa chegar ao fim do ano sem hesitações e que a aprendizagem possa ter tido, realmente, a maior produtividade possível.
2 – Erros do professor nunca devem repercutir-se na avaliação dos alunos. "É mais cego o que não quer ver, do que o que não vê". O professor nunca deve fazer «pagar» aos alunos erros que não são da responsabilidade deles, mas de si próprio.

Avaliação de programas e investigação

O processo de avaliação de que temos vindo a falar, consiste, portanto, em três partes fundamentais:

— Primeira: recolhem-se os dados para verificar a quantidade e qualidade da aprendizagem dos alunos;
— Segunda: tratam-se os dados para verificar a eficácia global do sistema ensino-aprendizagem;
— Terceira: julgam-se os resultados em função dos níveis considerados desejáveis para introduzir modificações futuras de modo a melhorá-los.

Muitas vezes se ouve dizer que, «os programas são grandes demais», «é impossível cumprir tal programa», «os programas não estão adequados ao nível etário dos alunos», «o programa não presta», etc. Tais afirmações escondem muitas vezes, verdades aparentemente óbvias e outras vezes «desculpas de mau pagador», sendo difícil apoiá-las ou contradizê-las por não existir avaliação de programas em Portugal. Esta avaliação que não se refere ao programa de estudo estritamente dito, mas também ao tipo e qualidade da sua implementação e ao desenvolvimento curricular efectuado a partir dele, é mais do que um problema de cada professor, um problema de cada instituição escolar e da comunidade escolar nacional no seu todo. É um vector fundamental da tal investigação pedagógica que pouco se faz, da tal inovação de que tanto se fala, e da tal avaliação de qualidade e mérito em que se pensa, mas nada se diz.

Não basta saber o nível de insucesso de um determinado ano ou ciclo (quando se sabe), é necessário investigar as razões do mesmo e actuar sobre aquelas em que seja possível introduzir modificações visando o melhoramento.

Em certas escolas, após o fim das actividades lectivas, ouvem-se, por vezes, os professores dizer que lhes foi marcado serviço de estatística. Isto é dito com o ar de quem tem, contra a sua vontade, de ir desempenhar mais uma tarefa burocrática que nada lhe diz. Ora, tal trabalho, não deve ser de modo nenhum somente um trabalho de estatística, mas sim um verdadeiro trabalho de investigação, visando a avaliação institucional e programática do ano findo. Saber quantos e quais (se necessário) os alunos que tiveram insucesso e em que disciplinas, quais os níveis de sucesso em cada uma destas, quais os resultados de cada professor e cada grupo, quais os níveis de cumprimento dos programas e as suas razões, qual a taxa de abandono, etc., etc., não serve somente para «saber». Serve para planear a introdução de mudanças que melhorem os resultados.

O objectivo de qualquer organização social tem que ser a melhoria constante do serviço que presta. Na educação é uma obrigação absolutamente prioritária.

O tempo que muitos professores passam a conferir bilhetes de identidade e colar selos no momento das matrículas, seria bem mais proveitoso quando utilizado em trabalho visando a investigação e avaliação de programas.

Já não estão em causa aspectos estruturais como a ausência de uma verdadeira carreira profissional motivadora, nem a pouca eficácia de um processo de regulação do sistema escolar, mas sim somente um pouco de organização durante as férias escolares, de forma a libertar os professores de tarefas administrativas (quando não de limpeza) e a enquadrá-los e organizá-los em equipas tendo em vista a avaliação e a investigação.

Não diremos que se passe já para um sistema em que cada professor seja um professor-investigador, mas, que, pelo menos, cada organismo, cada escola, avalie, no fim de cada ano, o trabalho realizado.

BIBLIOGRAFIA:

Bloom, B.S. (1976). *Human Characterististics and Shool Learning,* N.Y., U.S.A.: Holt, Rinehart and Winston.

De Bartolomeis, F. (1981). *Avaliação e Orientação — Objectivos, Instrumentos, Métodos,* Lisboa: Livros Horizonte.

De Landsheere, G. (1979). *Avaliação Contínua e Exames — Noções de Docimologia,* Coimbra: Livraria Almedina.

Dikic, M. (1986). A avaliação Formativa na Encruzilhada da Investigação e da Prática Pedagógica, in ALLAL et. al. (ed). *A Avaliação Formativa num Ensino Diferenciado,* Coimbra: Livraria Almedina.

Gagné, R. e Briggs, L. (1979).*Principles of Instructional Design,* 2nd ed., N.Y., U.S.A.: Holt, Rinehart and Winston.

Haayman, J. e Napier, R. (1979).*A Avaliação nas Escolas: Um Processo Humano para a Renovação,* Coimbra: Livraria Almedina

Joyce, B. e Weil, M. (1980), *Models of Teaching, 2nd ed.,* Englewood Cliffs, N.Y., U.S.A.: Prentice-Hall, Inc.

Sadler, D.R. (1983), Evaluation and Improvement of Learning. *Journal of Higher Education, 54,* 60-79

·EDUCAÇÃO HOJE·

O PATRIMÓNIO E A ESCOLA — *Do passado ao futuro — Isabel Cottinelli Telmo*

O COMPUTADOR — *Um instrumento de educação — João Ponte*

ENSINAR A LER, APRENDER A LER — *Ramiro Marques*

O CRITÉRIO DO SUCESSO — *Técnicas de avaliação da aprendizagem — Valter V. Lemos*

A FORMAÇÃO DE PROFESSORES — *À luz da Lei de Bases do Sistema Educativo — Manuel Ferreira Patrício*

A CRIATIVIDADE NO ENSINO DO PORTUGUÊS — *Ana M.ª R. Santos/M.ª José Balancho*

GOSTAR DE HISTÓRIA — *Um desafio pedagógico — M.ª do Céu Roldão*

DESENHO INFANTIL — *Um estudo sobre níveis do símbolo — M.ª Isabel Gândara*

A ESCOLA E OS PAIS — *Como colaborar? — Ramiro Marques*

EDUCAÇÃO AMBIENTAL — *Um guia prático para professores, monitores e animadores culturais e de tempos livres — Luís Filipe Oliveira*

A PUBLICAR

ACTIVIDADES PARA O DIRECTOR DE TURMA* — *Ramiro Marques*

ACTIVIDADES MATEMÁTICAS NA SALA DE AULA* — *Maria José Delgado/M.ª José Correia de Oliveira/Cristina Loureiro/Isabel Moura*

LEGISLAÇÃO ESCOLAR* — *João Bernardo*

* título provisório