

AVALIAÇÃO CONTÍNUA E EXAMES

NOÇÕES DE DOCIMOLOGIA

Gilbert DE LANDSHEERE



LIVRARIA ALMEDINA
COIMBRA-1976

**AVALIAÇÃO CONTÍNUA
E EXAMES**

NOÇÕES DE DOCIMOLOGIA

Título do original:

EVALUATION CONTINUE ET EXAMENS
— PRÉCIS DE DOCIMOLOGIE

Troisième édition revue et augmentée — 1974

© Edition Labor — Rue Royale, 342 1030 — Bruxelles

Tradução de

MANUEL PEREIRA DE CARVALHO

Todos os direitos reservados para a Língua Portuguesa
pela Livraria Almedina — Coimbra

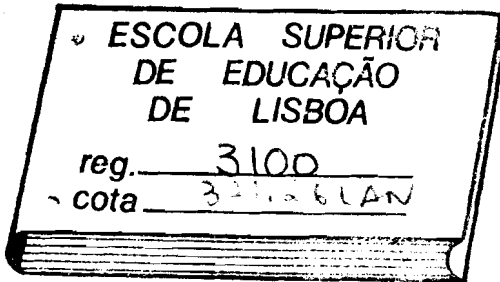
Gilbert DE LANDSHEERE

Professor da Universidade de Liège

AVALIAÇÃO CONTÍNUA E EXAMES

NOÇÕES DE DOCIMOLOGIA

Terceira edição revista e aumentada



LIVRARIA ALMEDINA
COIMBRA—1976

INTRODUÇÃO

Novas modalidades de avaliação dos alunos vão entrando nos nossos hábitos pedagógicos. Os resultados da *observação contínua* vão substituindo progressivamente o exame do fim do ano.

A inovação é louvável. Não traduz de modo algum o desejo de suprimir a *medida* na escola, mas, pelo contrário, de a tornar mais justa e cientificamente mais exacta.

Esta mudança ultrapassa de longe o domínio da atribuição de notas e coincide com transformações profundas da educação. O acontecimento não é fortuito. A civilização contemporânea, a economia das nossas sociedades reclamam um homem provido de novas qualidades e conhecimentos. Neste caso, impõem-se irresistivelmente as inovações pedagógicas.

A contestação dos exames tradicionais criou, no entanto, um mal-entendido grave. Assemelha-se de maneira surpreendente ao que surgiu, sobretudo entre 1920 e 1940, quando a adopção de ideias mal digeridas da pedagogia funcional dos Claparède, Dewey e outros «progressistas» conduziu ao culto da improvisação, ao desprezo da disciplina rigorosa, ao centrar da atenção em interesses que só existiam na cabeça dos teóricos...

Com o aparecimento de Dewey a escola poderia ter tomado como divisa: «Morreu o esforço, viva o esforço!». Hoje seríamos igualmente tentados a propor: «Os exames morrem, os exames morreram, vivam os exames!».

Não há dúvida que os velhos exames a envenenar a atmosfera e a matéria de ciclos inteiros de estudos e a decidir da

carreira escolar ou profissional em algumas horas — senão mesmo alguns minutos! — devem desaparecer. E não há dúvida também que a maior parte dos trabalhos docimológicos publicados até agora não têm sido mais do que a elaboração de sistemas aperfeiçoados para continuar a fazer mal as coisas. É que com efeito não se alicerçam sobre uma revisão completa do ensino.

Mas se por outro lado tivermos em conta, em todas as actividades do ensino básico, o lugar reservado à avaliação, ao *feedback* como dizem os psicólogos actuais, claro que os exames surgem de maneira *quase* permanente na vida escolar. Uma vez consistem em breves avaliações, outras tomam a forma de provas mais longas, abrangendo não raro quantidades consideráveis de matérias.

O essencial é que não venham inserir-se, como corpos estranhos, no processo de educação, mas façam dele parte integrante. Por esta mesma razão não se devem acantonar no domínio estreito do conhecimento mas sim visar a personalidade no seu conjunto.

Toda a reforma profunda é acompanhada, quase necessariamente, dum período de adaptação, de tateamentos, de erros. Assim se explica o desapontamento de muitos professores que estão a ensaiar actualmente o sistema de avaliação contínua.

Assim se explica também que caia, repentino, sobre a palavra exame, um tabu e que, no próprio instante em que escrevemos estas linhas, aqui e além, se vão passando certificados de fim de curso baseados em avaliações a partir de notas, qualitativa e quantitativamente insuficientes, integradas e baseadas para mais em impressões pessoais que nos levam a concluir que os velhos exames napoleónicos, comparados com elas, são monumentos de objectividade! Em todo o caso parece-nos que isto não passa duma crise de adaptação.

Não é, portanto, paradoxal que depois de termos recommençado — brevemente — o processo dos exames e de nos termos declarado partidário ardente da observação contínua e do balanço

de fim de ciclo, consagremos a presente obra... à organização dos exames. Mas o leitor sabe muito bem que, se a palavra permanece a mesma, o sentido mudou profundamente.

★
★ ★

Muitos dos docentes desconfiam ainda dos testes e de outros instrumentos que ambicionam quantificar o humano. Não sem razão, aliás, porque esses instrumentos manifestaram, nos últimos decênios, a sua doença de juventude. E como os seus utilizadores nem sempre souberam compensar a imperfeição das técnicas com uma grande moderação de juízos e com uma inteligência profunda das situações de conjunto, numerosos erros foram cometidos.

Mas se a prudência é sempre necessária, se o espírito crítico e o sentido clínico nunca perderão os seus direitos, foram realizados no entanto progressos consideráveis. Hoje podemos medir de modo satisfatório muitos comportamentos humanos.

Poderiam ainda alguns rejeitar a medida objectiva, não já porque lhe falte completa validade, mas porque não lhe compreendem a economia. Negar todo o valor ao que se ignora é um reflexo de defesa já bastante conhecido. Vem ainda complicar a situação o aspecto matemático dos métodos a adoptar e da literatura experimental, tantas vezes indigesta.

Os tempos de oposição entre psicometristas e práticos da educação já foram ultrapassados. Os primeiros deverão humanizar os seus números; os segundos, introduzir mais rigor nos seus processos. Ambos têm de unir e harmonizar os esforços para maior benefício do estudante e da comunidade.

★
★ ★

Acalmem-se os que, ao folhear este livro, encontraram alguns números. Não será precisa mais bagagem do que as quatro operações aritméticas.

Como é habitual, os processos que iremos considerar são mais difíceis de explicar que de aplicar.

Claro que nem todos os segredos da medição e da estatística serão revelados, mas as noções simples que vamos encontrar têm um valor prático comprovado. Além do mais, apresentam-se com um inegável valor de desmistificação e irão facilitar depois a leitura de trabalhos docimológicos mais especializados.

*
* *

O acolhimento dado às edições anteriores deste trabalho e a rápida evolução que sofrem os métodos de avaliação encorajam-nos a desenvolver e completar amplamente esta terceira edição. O essencial dessas novas achegas diz respeito à avaliação centrada nos objectivos e na metodologia que conduz ao domínio das aprendizagens.

PRIMEIRA PARTE

DEFINIÇÕES

I. DOCIMOLOGIA, DOCIMÁSTICA E DOXOLOGIA

A *docimologia* é uma ciência que tem por objecto o estudo sistemático dos exames, em particular do sistema de atribuição de notas e do comportamento dos examinadores e dos examinandos.

A *docimástica* é a técnica dos exames.

A princípio, a docimologia tomou um carácter negativo, criticando as maneiras de classificar e mostrando experimentalmente a falta de fidelidade e de validade dos exames.

Em seguida, entrou numa fase construtiva, tentando propor métodos e técnicas de medida mais objectivos ou, pelo menos, mais rigorosos e afirmando os meios de tornar as notas comparáveis, por forma a assegurar mais justiça escolar.

A *doxologia* é o estudo sistemático do papel que a avaliação desempenha na educação escolar.

J. Guillaumin¹ atribui-lhe, particularmente, os seguintes objectivos:

- estudo dos efeitos inibidores ou estimulantes das diferentes formas de exame;
- estudo das reacções emocionais dos alunos e, a partir daí, das reacções intelectuais aos juízos do professor;
- estudo dos efeitos de opinião do professor, no que toca aos alunos, sobre o seu ensino e sobre a aprendizagem escolar;

¹ J. GUILLAUMIN, «L'Aspect interpersonnel de la notation: de la docimologie à la doxologie pédagogique», in *Bulletin de la Société A. Binet et T. Simon*, 86, 1968, 250-275.

— estudo dos processos postos em jogo e dos efeitos obtidos pela automatização, pela interclassificação, pela classificação de equipa, pela ausência de classificação.

II. EXAMES E CONCURSOS

OBSERVAÇÃO E AVALIAÇÃO CONTÍNUAS

O exame e os concursos dificilmente podem dissociar-se da ideia de *prova* que se tornou aliás um sinónimo deles, depois de ter designado de forma mais corrente o sofrimento, a infelicidade, o perigo que revelam coragem e resistência.

No exame, a admissão é determinada por uma nota que o candidato deve atingir ou ultrapassar, ao passo que no concurso o número de lugares propostos é fixado de antemão. A presença da *ameaça*, do *perigo*, do *medo*, é inegável tanto num caso como no outro. Por menos imperfeições graves que esse procedimento apresente, não estamos muito longe do ordálio.

Na noção de concurso e exame, apercebemo-nos dum *carga agressiva* totalmente ausente dos conceitos de observação e de avaliação contínuas. A serenidade, a benevolência, bem como a indulgência, impregnam o professor que segue, com uma simpatia da qual nem toda a severidade é excluída, o longo encaminhamento dos alunos para o equilíbrio do momento e para o acesso pleno ao estado de adulto, mais tarde.

Os exames assinalam o fim das etapas; os concursos abrem as portas aos eleitos. São pontos na duração, acontecimentos no processo educativo.

A aprendizagem que precede estes acontecimentos consiste fundamentalmente numa sucessão contínua de comportamentos e de *feedbacks*, ou seja de informações a esclarecerem a sua validade, a sua pertinência. Não vem a propósito travar aqui uma longa discussão sobre as modalidades e os efeitos do

reforçar dos comportamentos. O importante é que, sem ele, a aprendizagem parece não poder produzir-se.

A avaliação, no sentido restrito que lhe damos no presente trabalho, merece pois um lugar importante no ensino de que faz parte integrante. Ela tem sempre relação, directa ou indirectamente, com o progresso, em extensão ou em qualidade, da aprendizagem.

A avaliação desempenha três papéis:

- 1.º Um papel de prognóstico: será o aluno provido de qualidades intelectuais e caracteriais, dos conhecimentos necessários para abordar uma matéria nova ou um ciclo de estudos superiores? Estará na classe onde normalmente devia estar? Responder a estas perguntas equivale a pre-dizer o êxito na etapa que vai começar.
- 2.º Um papel de apreciação:
 - a) controlo das aquisições;
 - b) avaliação do progresso, no caso de se comparar o aluno a si mesmo;
 - c) situação do aluno num momento dado:
 - * na sua turma ou grupo de trabalho?
 - * no conjunto das turmas paralelas duma mesma escola?
 - * em conjuntos mais vastos: cidade, concelho, província, nação?

Não se trata forçosamente de proceder a um exame ou um concurso, mas de fazer o apuramento, de determinar a posição relativa.

- 3.º Um papel de diagnóstico

Porque é que não se obteve uma aprendizagem perfeita? Quais as matérias ou técnicas que o estudante domina insuficientemente, quais os processos mentais em causa?

Os instrumentos necessários à avaliação, ao exame ou concurso nem sempre diferem, mas sim os modos de os utilizar.

E assim para evitar precisões verbais fastidiosas empregaremos, nas páginas que vão seguir-se, a palavra exame nos dois sentidos diferentes, sentidos esses que o contexto tornará sempre claros: exame propriamente dito e, de modo mais geral, todo o processo pedagógico que tem por objecto uma medição de aprendizagem ou de conhecimentos.

Exames internos e exames externos

Em sentido estrito, o *exame interno* em determinado ramo é organizado e classificado pelo professor que o ensinou e feito pelos alunos que receberam esse ensino, nos quadros da turma ou da escola.

Em sentido mais lato, chamam-se internos os exames organizados independentemente em cada escola, exista ou não uma coordenação ou uma unificação por ramo, níveis e secções.

Por *exames externos*, costumam designar-se as provas organizadas e classificadas por júris independentes das escolas, à escala local, regional ou nacional. As mais conhecidas dessas provas são as do bacharelato francês. Citemos ainda, na Bélgica, os exames cantonais, no fim do ensino primário e, na Inglaterra, até aos tempos mais recentes, o *Eleven + Examination* à entrada no ensino secundário.

III. MEDIDA E AVALIAÇÃO

Segundo J. P. Guilford, medir é «atribuir um número a um objecto ou a um acontecimento de acordo com uma regra logicamente aceitável»¹.

A medida exige pois:

- 1.º Que os objectos ou, melhor dizendo, as propriedades desses objectos sejam claramente definidas — tanto quanto

¹ J. P. GUILFORD e B. FRUCHTER, *Fundamental statistics in psychology and education*, New York, Mc Graw Hill, 1973, 5.ª ed., p. 19.

possível — mediante comportamentos ou características observáveis (definições operacionais);

2.º Que haja uma regra que indique como fazer corresponder um número a cada objecto.

Com toda a precisão, uma medida traduz-se necessariamente em algarismos, o que nem de longe é o caso da avaliação.

Ao opôr medida e avaliação, H. Taba escreve ¹: «O processo de medida é fundamentalmente descritivo, porque indica quantitativamente em que grau se possui determinada característica. A medida em educação concentra-se em geral em determinadas características específicas, delimitadas e bem definidas. A avaliação depende da medida, mas abrange um perfil mais vasto de características e *performances*».

Esta distinção, útil para tornar os conceitos precisos, não pode aceitar-se sem importantes reservas. H. Taba supõe implicitamente que toda a avaliação passa por uma quantificação rigorosa. Ora estamos muito longe disso, particularmente no domínio das atitudes e em todos os casos de grande complexidade. Quem visita uma exposição de pintura, classifica as obras do artista, e por conseguinte avalia-as, segundo o seu gosto, os seus critérios pessoais, sem ter que (ou mesmo poder) referir-se primeiramente a uma ou várias medidas. E é desta maneira que se avaliam muitos comportamentos e muitas obras humanas.

IV. NOTAS E «SCORES»

A distinção entre notas e «scores» prestaria, julgamos nós, um grande serviço à educação.

Num ditado, o aluno pode cometer um certo número de erros cuja existência pouco nos importa avaliar: ou existem

¹ H. TABA, *Curriculum development*, New York, Harcourt, Brace and World, 1962.

ou não existem. Todavia, o número de erros assinalados não tem em si mesmo qualquer sentido educativo. Dar cinco erros no famoso ditado composto por Mérimée¹ é prova de um extraordinário conhecimento dos arcanos da ortografia. Noutras condições, o mesmo resultado revela uma grave pobreza. É portanto necessária também uma informação relativista.

Por «score» designaremos os resultados objectivos obtidos mediante um teste ou qualquer outra forma de avaliação *por contagem ou desconto de pontos, segundo regras fixas*: número de erros num ditado, resultado num teste estandardizado.

Por nota, entenderemos uma apreciação sintética que traduza a avaliação de uma «performance» obtida no domínio da educação.

A nota pode ser objectiva ou subjectiva, mas é sempre relativa. Atribuir a nota A a um aluno cujo resultado obtido se situa a um determinado nível numa aferição nacional diz respeito à primeira categoria; pôr-lhe um *bom* na composição é situá-lo na segunda.

O verbo *classificar*, definido por Robert como «apreciar mediante uma observação ou uma nota numérica», é do uso corrente.

Para vários docimologistas franceses², *dar nota positiva* indica que se atribuiu um *score* por contagem de pontos; dar nota negativa indica uma dedução de pontos. Um ditado ou uma tradução são geralmente classificados por forma negativa; pelo contrário, em trabalhos de Ciências e Matemáticas, quem corrige soma os pontos atribuídos, à medida que são satisfeitas certas exigências ou certos critérios.

¹ Ver particularmente G. NOIZET e J. J. BONNIOL, *o. c.*

² *N. do Tr.* Trata-se de um ditado composto pelo escritor francês em que acumulara, por divertimento, um grande número de dificuldades ortográficas. Dizem que vários académicos cometeram um número respeitável de erros e que Napoleão não saiu da prova com menos de quarenta...

SEGUNDA PARTE

A ACUSAÇÃO E A DEFESA

CAPÍTULO 1

CRÍTICA AOS EXAMES

H. Piéron fez uma crítica pormenorizada dos exames tradicionais na sua excelente obra *Examens et Docimologie*¹. Todo o professor a deveria ter lido.

O presente capítulo contém dois tipos de observações. Um são breves referências a imperfeições sobejamente conhecidas, incansavelmente evidenciadas pelos docimologistas; outras dizem respeito a fenómenos não tão bem estudados: estereotipia, efeito de halo, efeito edípico da predição... Ums vezes, ~~está em causa apenas o exame; outras vezes, toda a pedagogia~~ em que ele se insere.

1. **Corpos estranhos na educação, ao serviço duma pedagogia ultrapassada**

Quer se trate de interrogatórios periódicos ou de exames trimestrais ou anuais, a avaliação reduz-se não raro a um controlo de retenção de conhecimentos, deixando inexplorados, não só os aspectos mais importantes da inteligência mas ainda quase todos os traços da personalidade que uma educação bem compreendida deveria cultivar.

Encontramo-nos, de facto, diante das sequelas dum sistema pedagógico em que, como observa T. Bramed, as lições servem

¹ Paris, P. U. F., 1963.

de fitas transportadoras de conhecimentos e de valores, seleccionados em função dum papel predeterminado a desempenhar numa sociedade não democrática.

Neste contexto, o exame constitui uma espécie de controlo de fabricação, de verificação da conformidade com o molde, com o padrão, numa palavra com especificações impostas pela autoridade.

O desabrochar da pessoa humana pouco lugar ocupa nessas preocupações, porque a educação não é concebida para ela, mas para o serviço dum regime.

As tristes consequências desta situação foram milhentas vezes analisadas. Em vez de ser útil ao aluno, de o informar funcionalmente do valor dos seus comportamentos em curso de aprendizagem, de fazer surgir uma melhor adaptação, de fornecer uma avaliação aceite pois que o estudante nela participa sincera e espontaneamente — o exame é tão mal acolhido como uma declaração de imposto que também não é certamente geradora de civismo.

Além de tudo o mais, a rejeição profunda duma tal actuação, ao combinar o seu efeito com a pobreza intelectual das perguntas feitas, conduz a um resultado fácil de imaginar: nos quinze dias que se seguem a um exame, não é raro que se esqueçam 80% das matérias factuais.

Deste modo, se exceptuarmos a função social desempenhada pelo exame, o fiasco é completo: nem se educou nem se instruiu.

2. Ansiedade e «stress»

Numa civilização em que o êxito escolar condiciona o êxito material e social, ao ponto de M. Young ter julgado seu dever denunciar os perigos da «meritocracia», o exame que decide da passagem de classe ou da obtenção do diploma é temido pela criança e pela família. Mesmo no decurso do ano, as

~~provas de controle~~ de conhecimentos e de aptidões são abordadas com tensão e apreensão, o que não é evidentemente a condição ideal, e desnatura profundamente o papel educativo da medição das aprendizagens.

Circunstância agravante: entre nós, a tradição exige que os exames finais abrangendo todos os conjuntos de disciplinas se realizem consecutivamente em alguns dias. Na altura duma sessão universitária, tive um ano o triste privilégio de ser interrogado sobre catorze disciplinas em dois dias consecutivos... No ensino secundário, produz-se o mesmo fenómeno de concentração por forma repetida ao longo do ano, quando o envio do boletim periódico aos pais suscita um fervilhar de interrogações.

Em ambos os casos, a acumulação de provas e a sua preparação, contrariadas muitas vezes pelo acrescentar de matérias novas até ao último momento — há professores que nem sempre respeitam os períodos de revisão ou não os exploram da melhor maneira — fatigam consideravelmente o corpo e o espírito.

Há quem pretenda que essas condições difíceis têm em si mesmo um valor educativo. A vida moderna dificilmente nos poupa ao «stress» e teremos de estar preparados para ele. De acordo, se as coisas se passarem claramente e sem equívocos.

Antes de organizar um exame, importa definir com clareza o seu objectivo. Trata-se de controlar a aquisição de conhecimentos? De avaliar a resistência ao «stress»? Ou de medir e capacidade de restituir e usar conhecimentos em situação da «stress»? Os três objectivos são diferentes e exigem provas diversas.

3. Desigualdade-injustiça

No nosso sistema escolar, cada um dos professores redige as perguntas de exame destinadas aos seus alunos. O princípio é excelente e num país como a Grã-Bretanha onde tradicional-

mente todos os exames-chave eram confiados a comissões exteriores à escola, tende-se a atribuir a cada docente a responsabilidade do controlo. No entanto, os métodos anglo-saxónicos apresentam mais rigor que os nossos, porque se tomam medidas precisas para assegurar a possibilidade de comparar exames na região ou até no país. Aliás, vamos consagrar uma parte deste volume aos processos utilizados para atingir esse objectivo.

A liberdade, quase total, concedida aos nossos professores conduz a situações injustas.

As nossas escolas constituem mundos isolados cujas populações apresentam características tantas vezes diferentes. Um certo estabelecimento de ensino secundário conta apenas com alguns alunos, na sua maioria desfavorecidos pela sua origem socio-económica. Ao contrário, num outro situado numa grande cidade, as secções mais importantes serão formadas por uma maioria favorecida social e intelectualmente, pois que os adolescentes em dificuldade poderão ser orientados para uma escola vizinha que goza da reputação de mais compreensiva.

Com o andar dos tempos, e com a ajuda da concorrência, vão-se criando nos estabelecimentos tradições de severidade ou de «generosidade». Marion Coulon observava recentemente que, quando se vão sucessivamente criando escolas de tipo idêntico numa mesma região, a severidade dos exames é inversamente proporcional à antiguidade do estabelecimento. Claro que não podemos generalizar, mas quem ousaria garantir que a afirmação é totalmente desprovida de fundamento? Já o compreenderam muitos pais que ambicionam para o seu filho um diploma determinado sem ligar muito ao valor intrínseco dos estudos. Em caso de insucesso, é mais fácil mudar de escola do que proceder a um tratamento ou modificar a orientação.

Seja como for, os professores adaptam o ensino e os exames ao nível da turma (o que é de louvar), resultando daí que um aluno bem classificado num conjunto fraco ficaria entre os fraquíssimos ou fracassaria num grupo bom.

O efeito cumulativo de todos os ajustamentos circunstanciais conduz por vezes a situações escandalosas. Há alguns anos pudemos consultar estatísticas que fixavam, para um mesmo tipo e um mesmo nível de ensino secundário, as possibilidades de êxito dum aluno que variavam de um a oito, segundo a escola frequentada. O caso é naturalmente extremo, e faltam-nos estudos precisos que permitam saber o que realmente se passa.

Convém insistir em que a adaptação do professor ao nível dos alunos não é, em si, digna de crítica; pelo contrário, é característica do bom educador.

A injustiça aparece quando notas puramente relativas são utilizadas em situação de concorrência exterior, ou são o único critério para a obtenção dum diploma de que pode depender o futuro dos alunos.

De qualquer forma, verificam-se variações consideráveis na *quantidade de matérias abrangidas* e na *qualidade das respostas exigidas*. Num caso, basta a reprodução memorizada do breve sílabo, ao passo que, no outro, as perguntas exigem análise, síntese e juízo pessoal em relação a uma matéria abundante.

Um determinado professor ~~faz apenas uma ou duas perguntas, escolhidas mais pela facilidade de correcção que oferecem do que pela sua importância real; outro, por seu turno, procura abranger todo o curso, sob risco de transformar o exame numa maratona.~~ No desporto, todos sabem que uma longa corrida deverá ser reservada a indivíduos excepcionais. E porque razão, no domínio do ensino, os trinta alunos duma turma deveriam ser corredores de fundo?

Antes de prosseguir, três resultados de investigações vão permitir-nos ilustrar os fenómenos em questão.

B. S. Bloom¹ comparou os resultados escolares de estudantes que terminavam o ensino secundário superior em 48 esta-

¹ B. S. BLOOM, «The 1955 Normative Study of the Test of General Education Development,» in *School Review*, 64, 1956, 110-124.

B. S. BLOOM and C. STATLER, «Changes in the States of the Tests of General Education Development from 1943 to 1955», in *School Review*, 65, 1957, 204-221.

dos diferentes dos Estados Unidos. O resultado médio entre o Estado com melhor classificação e o Estado com classificação inferior é de mais ou menos um desvio-padrão na distribuição normal dos *scores*. Este desvio corresponde de forma sumária a quatro anos de estudos.

O I. E. A. demonstrou recentemente (1973) que, no seio das nações industrializadas que participaram nas suas investigações, a diferença entre os *scores* médios dos países com resultados mais altos e os dos países com resultados mais baixos é da mesma ordem (matemática, ciências, literatura, leitura, inglês ou francês — língua estrangeira). Num país em vias de desenvolvimento, participando na investigação, e classificado em último lugar, a diferença eleva-se a dois desvios-padrões ou seja o equivalente a seis anos de estudo.

E Bloom conclui: «...segundo o lugar em que nasceu, o estudante deverá consagrar portanto de ano e meio a dois anos para atingir o nível que noutras zonas teria adquirido num ano»¹.

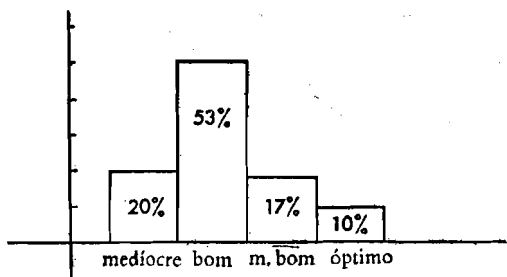
R. Gjorgjevski² mostrou, com uma experiência simples, como alguns professores avaliam diferentemente, por relatividade, uma mesma realização.

Cinco professores do mesmo ramo corrigiram e classificaram independentemente cem provas escritas provenientes duma mesma escola secundária. Em seguida escolheram-se quinze exemplares todos com a nota «*bom*». Para nova correcção, foram distribuídas a quatro outros professores. Estes adoptaram espontaneamente novas exigências. Só em cerca de metade dos casos, a nota continuou «*bom*». A nova classificação foi a seguinte³:

¹ B. S. BLOOM, *Time and Learning*, Comunicação ao 81.º Congresso da American Psychological Association, 1973.

² Ver N. ROT e Z. BUTAS, «Les distributions des notes scolaires comparées aux distributions des résultats obtenus aux tests de connaissances, in *Le travail humain*, XXII, 1-2, 1959.

³ O número de exemplares é pouco elevado para conduzir a resultados com significação estatística.



Proseguindo a experiência, Gjorgjevski extraiu da mesma série de cem exemplares um grupo de doze considerados muito bons e um grupo de doze considerados medíocres.

Em cada grupo, introduziu três exemplares considerados bons. Foram avaliadores duas vezes cinco professores. A média dos «bons» exemplares inseridos nos «muito bons» desceu de 3 em 5 para 2,40. No grupo oposto, a média passou de 3 a 3,87.

Devemos ainda a M. Reuchlin¹ um estudo sobre 4 860 alunos repartidos por 397 escolas primárias francesas. Os professores foram convidados a classificar os seus alunos em quatro categorias: muito bons, bons, médios e medíocres. Em seguida aplicou-se o mesmo teste de francês e de cálculo a todos os alunos. A verificação dos resultados mostrou que alunos que, segundo a escola, eram considerados muito bons, bons, médios e medíocres tinham o mesmo grau de conhecimento.

Como pertinentemente sublinha M. Reuchlin: «Pode-se dizer (...) que melhor que ninguém é o professor que conhece os pontos do programa assimilados ou não por cada um dos alunos. Mas o que ele ignora é a gravidade inerente a cada fraqueza, a cada lacuna, quando é considerada não em relação a uma turma que pode ser «boa» ou «fraca» mas ao conjunto do país. Daí as divergências de apreciação postas em foco pelo inquérito».

¹ Ver *Le Travail Humain*, XXII, 1-2, 1959, pp. 12 sg.

Ainda dois exemplos dados por F. Bacher¹. «Num inquérito francês que incide sobre uma amostragem representativa de alunos do terceiro ano (fim do primeiro ciclo secundário), foi possível fazer uma prova de conhecimentos literários e uma prova de conhecimentos matemáticos a alunos de 406 turmas. Verificou-se que as médias de turma se situavam entre 23 e 60 em francês (para uma prova classificada até 80) e entre 7 e 38 em matemática (para uma prova classificada até 44) (Reuchlin, Bacher, 1968). (...) Nos Estados Unidos, Flanagan (1964) faz notar que alunos classificados no quarto inferior da sua escola seriam classificados no quarto superior se frequentassem outras escolas da mesma região».

¹ F. BACHER, «La normalisation de la notation», in *Docimologie et Education*, número especial da revista *Les Sciences de l'Education*, 2-3, 1969, pp. 51-52.

Síntese dos resultados de M. Reuchlin (1959)

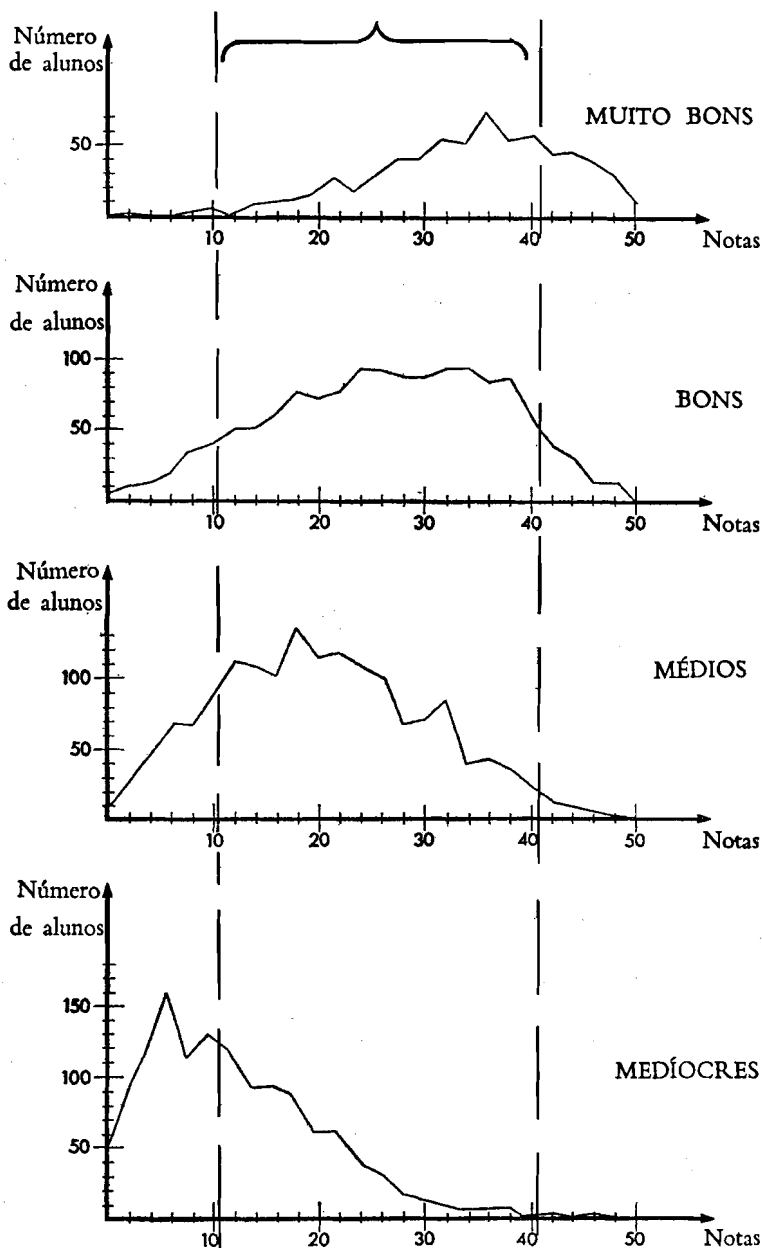


Fig. 1 — Os quatro gráficos permitem comparar a distribuição das notas obtidas, numa prova de aritmética, por 654 alunos considerados «muito bons» pelos seus professores, por 1303 alunos «bons», por 1551 «médios» e 1300 considerados «mediócras». Verifica-se que as distribuições de notas se recobrem largamente. Na zona de notas que vai de 10 a 40, o mesmo nível de conhecimentos pode ser qualificado, segundo os casos, com muito bom, médio ou medíocre.

4. O fracasso, gerador de fracassos

Pode-se objectar que os estudos superiores ou a vida irão encarregar-se de restabelecer a justiça que as notas escolares não tiverem respeitado. As coisas não são infelizmente tão simples. Em primeiro lugar, porque os alunos afastados ou desmoralizados injustamente no ensino secundário, por exemplo, não terão talvez ocasião de vir a defender a sua oportunidade no ensino superior. E depois os favorecidos poderão vir a ganhar em todas as situações.

As investigações actuais confirmam um velho adágio pedagógico: o êxito gera o êxito e o fracasso de hoje prepara o fracasso de amanhã. Impelidos pelo êxito, certos alunos valorizam um capital intelectual limitado. A passagem de classe «barata» tem às vezes a dupla consequência salutar de não levar os pais a retirar os filhos da escola em que se encontram e de permitir aos estudantes recobrem o entusiasmo. Quantas vezes não vêem os professores do ensino secundário subir espectacularmente o rendimento dum aluno desde o momento em que a crise psicológica da adolescência vai perdendo acuidade?

R. Rosenthal e L. Jacobson¹ publicaram recentemente um estudo de conjunto sobre o aspecto do *efeito edípico da predição*, assim designado porque se o oráculo não tivesse predito, no momento do nascimento de Édipo, que ele mataria o pai, nunca teria sido afastado da família. Conhecendo embora o pai, não o teria morto. Por outras palavras, a tragédia teve lugar, *porque fora predita.*

~~Parece que um aluno, em medida difícil de apreciar, se comporta em função da opinião que o professor tem dele.~~ Até os animais em domesticação estão sujeitos a esta regra. Rosenthal recorda a experiência seguinte. Formam-se dois grupos de ratos brancos geneticamente idênticos. No momento em que se entregam aos estudantes encarregados de os domes-

¹ *Pygmalion à l'école*, 1968.

ticar, indica-se que o primeiro grupo é composto de animais particularmente bem dotados, ao passo que o segundo é pobre em qualidade. Os resultados da domesticação confirmarão este prognóstico... fantasista.

A seguinte experiência, conduzida pelo próprio Rosenthal, situa-se no mesmo contexto.

Fez-se constar a professores dos seis anos de primária de Oak School que eminentes investigadores acabavam de criar um teste de «desabrochamento». Não passava realmente de um simples teste de inteligência ¹, pouco conhecido e sem virtude particular.

Foram submetidos a ele todos os alunos da escola e anotaram-se os que estavam na altura de desabrochar intelectualmente, prognóstico fantasista.

O quociente intelectual dos alunos assim designados elevou-se por forma significativa nos três anos inferiores. Além disso, foram observados progressos superiores à média em leitura e em aritmética ².

5. Rotura entre ensino e exame

~~O exame deve ser o reflexo do ensino feito~~ Imaginemos que um professor de educação física, em vez de levar os alunos à piscina, tivesse passado o ano a dissertar sobre a natação. Claro que seria inadmissível que as notas de exame fossem atribuídas de acordo com os resultados natatórios... porque só obteriam êxito os que tivessem aprendido a nadar fora do curso.

Parece inadmissível um tal procedimento. E, no entanto, quantas vezes não se assiste a uma aberração análoga. O pro-

¹ Medido com o auxílio do Toga de Flanagan (1960).

² Os números apresentados por Rosenthal e Jacobsen são contestados por R. L. Thorndike. Pelo contrário, a tendência à subida do Q. I. e dos «scores», parece reconhecida por todos os especialistas. É o que importa, evidentemente. Para mais pormenores, ver *American Educational Research Journal*, 5, 4, 1968, 708.

fessor faz um curso em que impõe as informações, as opiniões e os juízos; depois, no exame, faz as chamadas perguntas à inteligência que põem em jogo capacidades que não foram determinadamente preparadas durante as lições. Porque haviam os estudantes de ser capazes, dum momento para o outro, de descobrir sozinhos, no exame, a solução de aplicações geométricas originais, se não foram capazes de as encontrar durante o ano?

Assim se explicam muitas percentagens de fracassos anormais. Um inquérito originado por numerosas notas insuficientes dadas por um professor de química (4 alunos em 22 tinham obtido mais de metade dos pontos, no exame do fim do ano) revelou rapidamente que o professor fazia apenas dois ou três interrogatórios escritos durante o ano. Das três perguntas de exame, duas implicavam interpretação e *transfert*, operações que, no nível de dificuldade em que o professor se colocava, exigiam uma compreensão profunda dos fenómenos e um treino, longo e delicadamente controlado, na solução de problemas originais.

Uma discussão amigável com o jovem professor em questão revelou que ele se julgava obrigado a tratar em pormenor todos os pontos do programa, que não tinha consciência da necessidade de treino sistemático, e pelo menos semi-individualizado, em resolver problemas originais e que, fosse como fosse, não tinha as noções psicopedagógicas de base indispensáveis ao processamento dum treino dessa natureza.

6. Desacordo entre os que corrigem as provas

Todos os docimologistas citam exemplos de grandes divergências entre os professores chamados a avaliar um mesmo trabalho do aluno ou um mesmo conjunto de trabalhos, o que no entanto permite uma relatividade mais segura.

Todos sabemos como o nível de exigências varia segundo os examinadores. E aqueles que corrigem terão ao menos

uma consciência objectiva do seu grau de severidade ou de indulgência? Muito longe disso. R. Duquenne dividiu-os em dois grupos: os que se dizem severos e os que se dizem indulgentes. Foram corrigidos por todos, quatro exercícios. A média dos professores que se julgavam severos era de 12,4; a dos «indulgentes» era de 11,6.

As flutuações das notas dum mesmo professor, ao avaliar um mesmo trabalho passado um certo intervalo de tempo (fidelidade), podem ainda ser importantes.

Como todas as obras de docimologia estão pejadas de exemplos, só daremos alguns, como indicativo.

a) *A composição*

É sobre ela que caem as grandes acusações.

* Piéron refere que uma mesma composição avaliada por 76 elementos, todos professores de língua materna, conduziu aos seguintes resultados ¹:

<i>Nota na base de 20</i>	<i>Número de professores que corrigiram</i>
de 0 à 1	1
2 - 3	6
4 - 5	20
6 - 7	34
8 - 9	10
10 - 11	3
12 - 13	2

* O Serviço de Investigação pedagógica do C. R. D. P. de Lião convidou, em 1967 ², 150 professores de francês da

¹ PIERON, o. c., p. 123.

² C. R. D. P., *Docimologie et Examens*, Lyon, I. N. P., 1969.

Academia de Lião para corrigir três composições francesas que tratavam dum mesmo assunto.

Móx. 20	Médio	Margem de variação
Tema I	10 1/4	4 1/2 - 13 3/4
Tema II	6,5	2 1/2 - 12 1/2
Tema III	11,6	5 1/2 - 17 1/2

* A intervenção de homens cultos, estranhos às rotinas do ensino, melhoraria a situação?

J. French, P. Diderich e S. Carlton¹ compararam a maneira de corrigir de dez professores de língua materna, de nove professores de ciências naturais, de dez escritores ou redactores de jornais, de nove juristas e de sete directores de empresas comerciais.

A fim de obter uma larga distribuição dos resultados, foram convidadas várias centenas de estudantes do primeiro ano da universidade a tratar um dos dois temas seguintes: «Quem deve frequentar a universidade?» ou «A partir de que momento é que o adolescente deve ser tratado como adulto?» Recolheram-se cento e cinquenta trabalhos para cada um dos dois temas. Segundo o cálculo das probabilidades, cada amostragem deveria incluir normalmente exercícios muito bons e muito maus, o que tornava a discriminação tão fácil quanto possível.

Os trezentos trabalhos foram avaliados por todos os professores que trabalhavam cada um para si em casa e apenas tinham recebido as seguintes indicações:

1.º Tenha em conta a sua própria maneira de ver para definir o que pensa ser a «faculdade de escrever»;

¹ J. W. FRENCH, *Schools of thought in judging excellence of English themes*, Princetown, Educational Testing Service, 1961.

- 2.º Distribua os trabalhos em nove séries, por ordem de mérito;
- 3.º Procure que pelo menos seis das 150 redacções tratando do mesmo assunto figurem em cada série;
- 4.º Mediante um comentário em cada trabalho, indique a razão porque gosta dele ou porque não gosta.

Calculou-se a correlação entre todos os pares possíveis de professores a corrigir. A correlação média é de 31.

Em 300 trabalhos:

- 101 obtiveram todas as nove classificações possíveis;
- nenhum trabalho teve menos de cinco classificações diferentes.

b) *Matemática*

O C. R. D. P. de Lião¹ pediu também a 150 professores de matemática que davam aulas numa turma do 5.º ano para classificarem os trabalhos de 3 alunos que tinham de resolver o mesmo problema.

Máx. 20	Média	Margem de variação
Aluno I	5,70	0,5 - 11,5
Aluno II	16	11,5 - 20
Aluno III	8	3,5 - 11,5

Se para o aluno I, por exemplo, se eliminarem as duas notas extremas, serão tantos os professores a dar a classificação de 2/20 como de 8/20, ao passo que se esperaria uma classificação objectiva.

c) *Medicina*

Ao nível universitário, os resultados não são mais seguros.

¹ O. c.

Um mesmo aluno do 2.º ano de Medicina teve, em Junho e em Setembro, de 1965, uma prova escrita com cinco perguntas simples e precisas a tratar em 1 h. 30¹.

As respostas (anónimas) foram corrigidas independentemente por cinco professores com cursos e qualidades idênticas.

Resultado total em 100	Margem de Variação	Decisão	
		Admitido	Reprovado
Exercício de Junho	47 - 78	3	2
Exercício de Setembro	32,5 - 73	3	2

Se considerarmos as notas atribuídas a cada uma das 10 perguntas, verifica-se:

1) desvio máximo	12	(de 3 a 15 na escala de 20)
2) desvio para 4 perguntas	11	
3) Menor desvio máximo entre 5 avaliadores	7,5	

d) *Diversos*

A. Agazzi² recorda um exemplo mais impressionante ainda, relativo a um conjunto de secções (nota de reprovação: menos de 10 em 20):

	Rejeitados pelos seis avaliadores	Admitidos. pelos seis avaliadores.	Admitidos por uns e recusados por outros.
Tradução latina	40 %	10 %	50 %
Composição Francês	21 %	9 %	70 %
Inglês	37 %	16 %	47 %
Matemática	44 %	20 %	36 %
Filosofia	9 %	10 %	81 %
Física	37 %	13 %	50 %

¹ Ver *Le Monde* de 6 de Setembro de 1966. Note-se que os três admitidos foram os mesmos nos dois casos.

² A. AGAZZI, *Les aspects pédagogiques des examens*, Strasbourg, Conseil de l'Europe, 1967, p. 119. O estudo é de H. Laugier e D. Weinberg e é relativo ao bacharelato francês.

e) *Nas provas orais, mais discordâncias ainda*

Todos os docimologistas estão de acordo para reconhecer mais discordâncias ainda no caso dos interrogatórios orais do que dos interrogatórios escritos. Raras são, no entanto, as experiências rigorosamente controladas nesta matéria. Uma das mais recentes foi realizada por H. Piéron, M. Reuchlin e F. Bacher ¹.

f) *Quantas pessoas a corrigir para estabilizar a nota?*

Traria uma melhoria considerável calcular a média entre dois avaliadores? Não traz, a não ser que tenham sido tomadas medidas precisas para os pôr de acordo. Por quantos professores diferentes conviria avaliar o mesmo trabalho para estabilizar a nota? Os números seguintes fornecidos por Piéron ² não têm evidentemente valor absoluto. Dão todavia uma ideia da ordem de grandeza dos números...

Composição	78
Versão latina	19
Inglês	28
Matemática	13
Dissertação filosófica	127
Física	16

A história e a geografia não figuram na lista. Mas esses dois ramos dão também origem a grandes flutuações.

Impõe-se finalmente uma observação: estabilizar a nota em particulares circunstâncias de exame não implica de maneira nenhuma que se tenha obtido *a verdadeira nota*. É evidente que nem um indivíduo que executa uma tarefa nem a mesma tarefa em si contêm uma nota que represente o seu valor essencial, absoluto. «Esta noção, este mito da verdadeira nota merecida por um exercício, aparece à análise enganador sob o plano teórico, abusivo sob o aspecto metodológico e

¹ H. PIÉRON, M. REUCLIN e F. BACHER — *Une recherche expérimentale de docimologie sur les examens oraux de physique à une session du baccalauréat*, in *Biotypologie*, 1962, 23, 48-73.

² O. c.

absurdo sob o plano prático. No entanto, apresenta-se cómodo e ideologicamente orientado para perpetuar a ideia desprovida de todo o fundamento de que existiria em qualquer parte, na harmonia preestabelecida, uma nota que seria a justa medida dum exercício. E isto implicaria que essa nota viria a ser «a nota merecida por aquele que a recebe»¹.

Noizet e Bonniol² chegam a duvidar mesmo da possibilidade de estabilização duma nota. Segundo eles, de modo nenhum estaria assente que as causas de flutuação estivessem aleatoriamente distribuídas entre os que corrigem e portanto se compensam. Laugier e Weinberg, seguidos por H. Piéron, admitem francamente esta hipótese. Nós partilhamos da sua opinião.

Poderiam multiplicar-se os exemplos de desacordo entre os que classificam. Mas que interesse tem o insistir na acusação fácil, como têm feito muitos docimologistas? O importante é ter a consciência do perigo e, sobretudo, procurar os remédios.

Qual é a explicação das divergências observadas?

Reside fundamentalmente na diversidade de pontos de vista, das dimensões segundo as quais uma mesma prova pode ser avaliada e da ausência de orientações precisas (por vezes absolutamente nulas) dadas para a classificação. F. Bacher³ afirma:

«Numa dissertação podemos ter em conta a organização das ideias, a originalidade, a correcção da expressão, a elegância do estilo, etc. Se considerarmos cada aspecto possível como uma dimensão de classificação, o trabalho dum aluno pode situar-se em *n* dimensões. A dificuldade nasce de que, nesse espaço, se estabelece uma única ordem parcial entre os alunos. Um aluno que ocupe uma posição elevada em todas as dimen-

¹ J. J. BONNIOL, *Les comportements d'estimation dans une tâche d'évaluation d'épreuves scolaires*, Aix-en-Provence, Université de Provence, 1972 (tese do 3.º ciclo).

² G. NOIZET e J. J. BONNIOL, *o. c.*, p. 784.

³ F. BACHER, *La normalisation des notes*, *o. c.*, pp. 53-54.

sões pode ser declarado superior a um aluno que ocupe uma posição menos elevada em todas as dimensões; mas não pode afirmar-se que um aluno que ocupa uma posição elevada numa primeira dimensão e uma posição fraca numa segunda dimensão é superior ou inferior a um aluno que ocupa posições inversas».

Além do mais, mesmo que considerem o mesmo aspecto, os avaliadores variam em severidade, em estabilidade de critério, em resistência ao efeito de halo, etc.

7. Infidelidade no mesmo avaliador

Um mesmo avaliador será pelo menos igual a si mesmo? Não. A fenómenos tão evidentes como as variações de saúde física e mental e como a evolução do saber, junta-se um grande número de factores mais ou menos bem definidos: variação na qualidade da relação estabelecida com o aluno, no contexto da avaliação (se acabámos de corrigir um trabalho excelente, o seguinte pode ser prejudicado, ...), na escala consciente ou inconscientemente adoptada, etc.

Foram convidados catorze historiadores para classificar pela segunda vez quinze composições, doze a dezanove meses depois de as terem classificado uma primeira vez. Tinha-se eliminado todo e qualquer traço de correcção. Os professores davam não apenas pontos mas indicavam o êxito global ou o fracasso.

De 220 casos, em 92 o veredicto foi diferente da primeira para a segunda vez ¹.

Convém no entanto insistir sobre o facto de que resultados tão pobres serão devidos à falta de orientações rigorosas que indiquem os aspectos dignos de consideração pelos professores que classificam.

¹ HARTOG e RHODES, *An Examination of Examinations*, London, McMillan, 1935, pp. 81 e 15.

Um esquema para continuar a investigação

É útil que, de tempos a tempos, se façam investigações sobre a fidelidade dum professor em relação a si mesmo ou sobre a concordância entre vários professores, quando mais não seja para reforçar a consciência de um perigo iminente. E é em virtude disso que julgamos oportuno assinalar o excelente plano de investigação devido a dois dos melhores docimologistas contemporâneos: F. Yates e D. Pidgeon.

- * Formaram-se 50 grupos de crianças que acabavam a escola primária. Foram submetidos a duas provas de língua materna:
 - a) redacção sobre um assunto escolhido de entre três ou quatro;
 - b) perguntas de compreensão de textos e perguntas de gramática (uso da língua).
- * Sete examinadores experimentados — cinco homens e duas senhoras — classificaram os trabalhos.
- * Primeiramente cada grupo de crianças fez três exames diferentes, com uma semana de intervalo; em seguida, cada uma das três provas foi repetida duas vezes com uma semana de intervalo.

ESQUEMA ¹

GRUPO	Sessões de exame				
	I	II	III	IV	V
1	Aae	Bac	Dab	Aac	Aae
2	Bbf	Cbd	Ebc	Bbf	Bbf
3	Ccg	Dce	Fcd	Ccg	Ccg
4	Dda	Edf	Gde	Dda	Dda
5	Eeb	Feg	Aef	Eeb	Eeb
6	Ffc	Gfa	Bfg	Ffc	Ffc
7	Ggd	Agb	Cga	Ggd	Ggd

¹ Feito por D. FINNEY, citado por Yates e Pidgeon, *Admission to Grammar Schools*, Londres, N. F. E. R., 1957, p. 99.

Sete séries de perguntas: A, B, C, D, E, F, G.

Sete examinadores: a, b, c, d, e, f, g.

Não vamos retomar o pormenor dos resultados, mas pretendemos simplesmente dar um exemplo de controlo bem conduzido, nomeadamente com uma medição da aprendizagem.

8. Estereótipos e efeitos de halo

Na situação escolar habitual, o professor conhece cada um dos seus alunos e pode assim dosear, matizar as notas em função dum efeito desejado. Num caso, encorajam-se, exagerando o valor do trabalho; noutra caso, toma-se uma atitude de severidade excepcional para efeito de choque que se prevê salutar. Em casos destes, o professor age deliberadamente, com plena consciência.

Mas já é totalmente diferente quando intervêm os fenómenos de estereotipia e de halo.

a) Estereotipia

Entendemos por *estereotipia* uma imutabilidade mais ou menos acentuada que se instala no juízo feito sobre o aluno.

A estereotipia resulta duma contaminação dos resultados. Um primeiro trabalho medíocre inclina a pensar que o segundo não o será menos; e, se isso se verificar, aumenta ainda a tendência a dar uma nota medíocre ao terceiro e assim por diante. No professor sobrecarregado de correcções, a deformação produz-se muito mais facilmente.

Seria errado julgar-se que a estereotipia influencia unicamente as avaliações de base subjectiva acentuada (dissertação, prova de história e, em geral, respostas que impliquem um juízo de valor). Atinge igualmente exercícios tão «objectivos» como o ditado de ortografia. Testemunha-o a experiência seguinte.

Um professor de língua materna faz ditados regularmente. Não tarda em conhecer os alunos que obtêm habitualmente bons resultados e os que mais facilmente falham. Se determinarmos os erros «esquecidos», não notados pelo professor, logo verificamos que os esquecidos referentes aos bons alunos são significativamente mais elevados do que os que dizem respeito aos alunos fracos. No primeiro caso, o professor não espera encontrar erros; no segundo, espiolha-os¹.

Se um caso de falta de consciência profissional intervier, a situação pode tornar-se muito grave. Tinham-nos chamado a atenção para um caso de estereotipia manifesta de que era vítima um aluno na tradução de latim (ensino secundário). A título de verificação, mandámos fazer os exercícios ulteriores sucessivamente a um outro aluno da turma, a um aluno do mesmo nível pedagógico, o mais classificado de outra escola, em seguida a um licenciado em Filologia Clássica. Não houve variação de meio valor na escala de vinte...

Repetimos este género de experiências noutras ocasiões. E quantos pais também, às vezes muito competentes no domínio em que faziam ocasionalmente o trabalho do seu filho, não sentiram uma ponta de decepção ou surpresa quando tiveram conhecimento da «sua» nota.

Concluir daqui que todos os professores incriminados assim pecam por falta de sentido das responsabilidades seria naturalmente injusto. São mais os métodos de avaliação do que os homens que deveremos pôr em causa.

b) *Efeito de halo*

O *efeito de halo* apresenta um carácter afectivo acentuado. Valorizam-se não raro as respostas dum aluno com boa figura, de olhar franco, de dicção agradável. Não devemos contudo

¹ Ver a este propósito M. ZILIG, «Beliebte und unbeliebte Volkshüterinnen», in *Arch. f. d. ges. Psychologie*, 12, 1934, p. 32, citado por E. HOHN, *Der schlechte Schüler*, Munique, R. Piper, 1967, p. 105.

generalizar. Conhecem-se professores que, por anticonformismo, favorecem um certo desmazelo ou originalidade que nem sempre são de bom gosto.

Ou por razões de legibilidade ou por razões claramente afectivas, a caligrafia pode também influenciar o professor que corrige. Os especialistas da publicidade sabem de longa data que a apresentação da mensagem tem uma considerável influência no seu rendimento.

C. Chase¹ estudou a influência da qualidade da caligrafia nas notas atribuídas às redacções. É perfeitamente significativa.

Verifica-se que, mesmo nos testes objectivos que exigem indicações manuscritas da parte do estudante e para cuja correcção se utilizaram escalas standardizadas, a má qualidade da caligrafia faz baixar o score.

As seguintes experiências devidas a R. Weiss² mostram com que facilidade pode ser provocada uma combinação de estereotipia e de efeito de halo.

R. Weiss seleccionou duas redacções feitas por alunos da 4.ª classe da primária. Os trabalhos foram dactilografados sem a mínima modificação, depois entregues para correcção a dois grupos de 46 professores da mesma 4.ª classe da primária.

Acompanhava a distribuição dos trabalhos ao primeiro grupo o seguinte comentário: «São duas composições escritas por alunos da 4.ª classe da primária. O trabalho n.º 1 foi feito por um aluno médio que gosta de ler as bandas desenhadas; o pai e a mãe são empregados. O trabalho n.º 2 foi feito por uma criança bem dotada; o pai é redactor dum diário conhecido». Para o segundo grupo de professores inverteram-se os papéis.

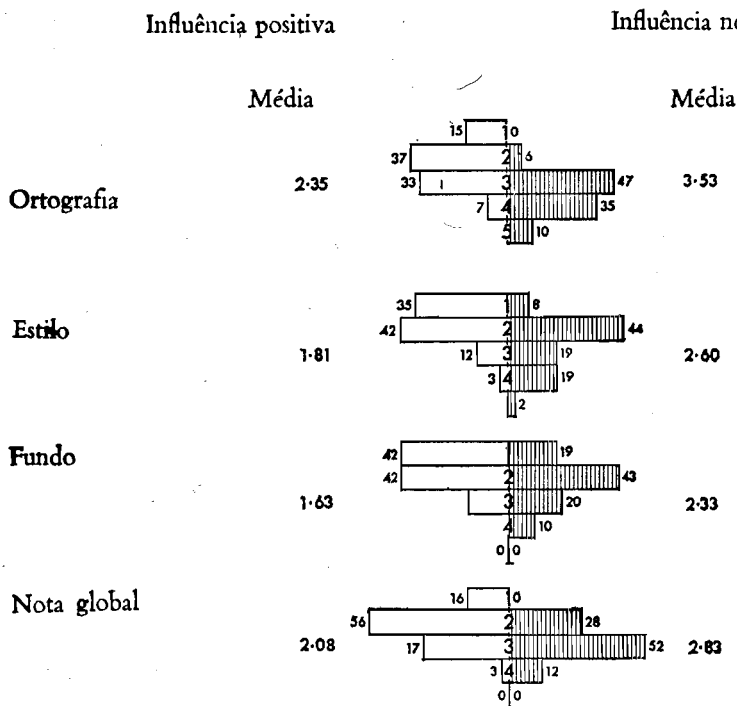
¹ C. CHASE, «The impact of some obvious variables on essay test scores», in *Journal of Educational Measurement*, 1968, 5, pp. 315-318.

² R. WEISS, *The Reliability of the Number Marking System: An Austrian Study*, in J. LAUWERYS e D. SCALON, *Examinations*, Londres, Evans Brothres, Ltd., 1969, pp. 101-107.

A correcção deveria ser feita tendo em conta uma escala de cinco graus: muito bom (1); bom (2); médio (3); suficiente (4) e insuficiente (5). A ortografia, o estilo, o fundo, depois o conjunto deveriam ser classificados separadamente.

Para os quatro aspectos considerados, as notas atribuídas ao trabalho a que se tinha criado um preconceito favorável foram significativamente superiores às outras. Para a ortografia, que se afigurava dever escapar mais ao efeito de halo, observa-se que para o aluno apresentado como dotado, 16% dos professores atribuem a nota *muito bom* e nenhum deles a nota *insuficiente*. Se o aluno foi apresentado como médio, nenhum deles dará *muito bom*, e 11% dará *insuficiente*.

NOTA DAS COMPOSIÇÕES SOB INFLUÊNCIA DE UM PRECONCEITO FAVORÁVEL OU DESFAVORÁVEL ¹



Numa experiência similar com problemas de aritmética, igualmente na 4.^a classe da primária, R. Weiss observa:

Preconceito favorável

- 11% muito bom
- 5% suficiente
- 0% insuficiente

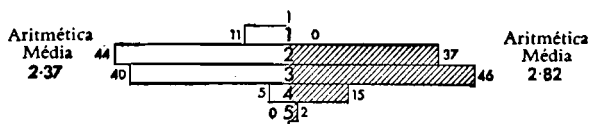
Preconceito desfavorável

- Nenhum muito bom
- 15% suficiente
- 2% insuficiente

PROBLEMA DE ARITMÉTICA

Influência positiva

Influência negativa



¹ Todas as diferenças são estatisticamente significativas de P.01, excepto uma (P.10).

Vê-se bem como um artifício grosseiro exerce já uma influência. Como sublinha Weiss, é bem maior provavelmente a deformação numa situação real em que o professor conhece o aspecto do aluno, o comportamento, a maneira de se exprimir, a profissão e a vida social dos pais, etc.

9. Efeitos da ordem de correcção

Os alunos batidos em exames há muito tempo que desco-briram a importância dos contrastes. O facto de estar a seguir imediatamente a um candidato muito brilhante pode ser prejudicial; vir depois de um mais fraco pode ser vantajoso, caso o professor que corrige não esteja mal humorado pela mediocridade das respostas. Numa palavra, a ordem de avaliação ou de correcção tem importância.

Este fenómeno foi estudado particularmente por J. J. Bonniol¹ que faz dele um dos determinantes essenciais da avaliação.

Numa primeira experiência, Bonniol constitui dois grupos de nove professores. O primeiro recebe uma série de exercícios de acordo com uma ordem dada e o segundo numa ordem inversa. Bonniol observa que as divergências (importantes) entre os dois grupos «são mais imputáveis às duas ordens de correcção que às diferenças de critério dos examinadores». Os examinadores classificam por contraste com o exercício precedente. São mais severos em relação aos últimos trabalhos classificados do que aos primeiros.

Para as explicar, J. J. Bonniol trata os efeitos de ordem de correcção como um problema de *percepção* dos estímulos numa série. «Como no caso de percepção o que se pede ao sujeito que deverá classificar uma série de exercícios é a organização de um código, o pôr em correspondência elementos de dois conjuntos: o das notas que lhe é possível utilizar e o das dimensões em função das quais podem ser avaliados os estímulos».

¹ J. J. BONNIOL, «Les comportements d'estimation dans une tâche d'évaluation d'épreuves scolaires». Étude de quelques-uns de leurs déterminants. Aix-en-Provence. Université de Provence, 1972 (nota de apresentação numa tese do 3.º ciclo).

Para verificar experimentalmente se os examinadores julgam bem por contraste, Bonniol introduz um ou vários exercícios muito bons ou muito maus numa série de exercícios médios. Os efeitos de ordem podem assim ser considerados como efeitos de ancoragem¹.

Surgem duas hipóteses fundamentais. A introdução das âncoras exercerá efeitos de contraste que se traduzem por *deslocações* na escala por sobreavaliação ou subavaliação dos trabalhos que se seguem à âncora na série, ou por modificações da *extensão* da escala utilizada.

Intervêm na experiência duas vezes sete grupos equivalentes de 17 examinadores.

Bonniol forma duas séries de exercícios médios: 17 traduções em inglês (notas negativas) e 14 soluções duma operação logarítmica simples (notas positivas).

Após a classificação destas séries médias, Bonniol introduz:

- uma âncora alta (exercício muito bom) no fim do primeiro terço da série;
- uma âncora alta (exercício muito bom) no princípio do terceiro terço da série;
- três âncoras altas (exercícios muito bons) no fim do terceiro terço da série;
- três âncoras altas (exercícios muito bons) no princípio do terceiro terço da série;
- uma âncora baixa (exercício muito mau), etc.
- três âncoras baixas, etc.

Para as traduções de inglês, a âncora exerce um efeito sobre um ou vários exercícios que se seguem. Estes efeitos são mais importantes com uma âncora pesada (três exercícios).

¹ Bonniol define a âncora como «um estímulo privilegiado que serve de estímulo de referência, ou porque está presente com mais frequência que os outros ou porque está situado numa posição particular ou porque duma ou doutra maneira é assinalado à atenção do sujeito». A introdução de âncoras vem modificar o modelo, o nível de *performance* a que o examinador se referia para avaliar.

A ancoragem baixa exerce maior efeito que a ancoragem alta (notas negativas).

Em matemática, é necessária a âncora pesada para exercer um efeito que se torna mais importante se as âncoras são altas (notas positivas).

10. Falta de validade

Os exames tradicionais não são apenas fracos meios de inventário; o seu valor prognóstico também se revela contestável em muitos casos.

Observámos, em companhia do reitor dum dos maiores liceus da Bélgica, que duas categorias de alunos obtêm resultados brilhantes no ensino universitário: os que, no decurso de todos os seus estudos brincaram com a dificuldade e permaneceram à cabeça da turma e uma parte não insignificante dos que quase nunca tiveram fracassos, mas no entanto se foram sempre mantendo no limite da nota fatídica. Em ambos os casos, as «performances» obtidas no ensino secundário têm real significado, não tanto pela sua natureza aparente como também pela facilidade de adaptação, pela plasticidade intelectual para esse efeito indispensável.

Prognosticar o êxito ou o fracasso na universidade, partindo da percentagem obtida no termo do ensino secundário é, portanto, muito arriscado.

11. Um instrumento de imobilismo social

O exame tradicional justifica-se quando mais não seja como meio de selecção e como instrumento de promoção da educação individual. Além disso, essa selecção opera-se, em muitos casos, menos em função da qualificação técnica que da qualificação social.

O exame torna-se assim um instrumento privilegiado do imobilismo social, fenómeno analisado de maneira penetrante

por J.-C. Passeron a quem vamos buscar o essencial das considerações que seguem¹.

a) *Efeitos irreversíveis da certidão escolar*

Uma nota de curso, uma vez obtida, segue o indivíduo por toda a vida, assegura-lhe grosso modo o seu grau na hierarquia profissional, o seu nível de remuneração e o poder de que irá dispor.

Ora, no momento da obtenção do diploma, «a competência é medida não através da actividade que será exigida ao trabalhador, mas através duma espécie de actividade analógica, preparatória, lúdica e quase fictícia. O que os exames medem em relação às exigências profissionais não é o que se julga que as pessoas serão capazes de fazer, mas antes a que nível será socialmente exigível pagar-lhes» (p. 7).

Claro que estas observações não têm valor absoluto. Correspondem, todavia, a uma realidade incontestável. Nunca se deu, por exemplo, o grau de engenheiro a um operário ou a um técnico que adquiriu a mesma competência profissional.

O último exame feito com bons resultados fixa muitas vezes um indivíduo numa zona socio-profissional de que é, ainda hoje, muito difícil sair. A precocidade desta fixação está em contradição profunda com toda a política contemporânea da educação e, portanto, da promoção permanente.

Que o exame dê mais uma certidão social do que uma **garantia** de competência técnica, lá estaria a prová-lo o facto de em muitas empresas duas pessoas ocuparem o mesmo posto:

«uma oficialmente, porque tem todos os diplomas escolares requeridos para ocupar (com remuneração, prestígio e poder), outra efectivamente, porque assume a parte técnica do trabalho

¹ J.-C. PASSERON, «Sociologie des examens», in *Education et Gestion*, 1970, 2, 6-16.

(...), isto é, faz «todo o trabalho» (...). Às vezes a segunda conviria tecnicamente ao posto, muito mais que o titular. Mas é precisamente isso que o efeito do certificado resultante do exame impede. Duma vez para sempre e para toda a vida, tende a definir a qualidade da prestação profissional dum indivíduo pelo valor desse indivíduo no mercado dos títulos escolares».

Mais uma vez esta verificação exige restrições. Numa empresa raramente basta aos quadros apenas a competência técnica. A capacidade de comunicar, de exercer funções de *leader* e de relações públicas, de harmonizar o trabalho da administração e da produção, etc., são tantas outras qualidades que escapam aos que vêem apenas o lado técnico das coisas. Mas essas qualidades não são sempre reais, muito longe disso. Não raro deparamos com um formalismo de que falaremos ainda.

b) *Os exames não são socialmente neutros*

Um inquérito do INSEE, em França, revela que, se as possibilidades de acesso ao ensino superior aumentaram para todas as categorias sociais durante os últimos anos, «o acréscimo global da taxa de escolarização processou-se entre as diversas categorias socio-profissionais na mesma proporção que as desigualdades anteriores. Numa palavra trata-se duma simples *translacção para o alto* da estrutura das desigualdades» (p. 8). O seguinte quadro é revelador.

<i>Categoria socio-profissional do pai</i>	<i>Probabilidades de acesso ao ensino superior</i>	
	1961-62	1965-66
Assalariados agrícolas	1,1	2,7
Agricultores	3,4	8
Operários	1,3	3,4
Empregados	9	16,2
Quadros médios	24,9	35,4
Profissões liberais e quadros superiores	38	58,7

Como se explica este fenómeno na aparência contraditório com a obrigatoriedade escolar generalizada e cada vez mais longa, com a gratuidade do ensino e a repartição das aptidões segundo as leis do acaso, em todas as camadas da sociedade?

A diferença de educação familiar criaria rapidamente, no decurso dos primeiros anos, diferenças nos mecanismos intelectuais, nos meios linguísticos, nas aptidões em face do trabalho: «A educação dada pelas famílias predispõe tanto menos ao êxito nos exames quanto mais afastada estiver a classe a que essas famílias pertencem da cultura escolar, da cultura mais vasta» (p. 11).

Segundo a família, o meio rural ou urbano, o nível de desenvolvimento do país em que nascem, indivíduos dotados dum mesmo potencial à partida têm probabilidades de êxito muito diversas.

O conjunto impressionante das últimas investigações da Associação Internacional para a Avaliação do Rendimento Escolar (I. A. E.)¹ confirma que, nas condições actuais, a origem familiar permite prever melhor a *performance* escolar do que qualquer outra variável.

B. Bloom², numa comunicação recente, indicou igualmente que, para igual tempo de estudos — 12 anos por exemplo —, o rendimento escolar médio pode variar do simples ao dobro, conforme o grau de industrialização do país.

Numa palavra, para voltar a Passeron, os exames teoricamente neutros, limitar-se-iam a registar os efeitos profundos da primeira educação. Teremos de reconhecer que esta tese se verifica muitas vezes. Não se trata contudo de um determinismo inevitável. Demonstram-no investigações actuais sobre

¹ L. C. COMBER and J. KEEVES, *Science Education in Nineteen Countries*, I. E. A., Stockholm, Malmqvist, 1973.

A. C. PURVES, *Literature Education in Ten Countries*, J. E. A., Stockholm, Malmqvist, 1973.

R. L. THORNDIKE, *Reading Comprehension in Fifteen Countries*, I. E. A., Stockholm, Malmqvist, 1973.

² B. S. BLOOM, *Time and Learning*, Communication au 81^o Congrès de l'American Psychological Association, Montréal, 1973.

a compensação precoce das deficiências socio-culturais¹. Mas resta incontestável que os exames actuais não são socialmente neutros.

«Os sistemas de classificação, continua Passeron, e os tipos de provas utilizados têm em conta pelo menos tanto aptidões técnicas como certos aspectos gratuitos da «performance» que não têm qualquer importância técnica mas que não deixam de estar fortissimamente ligados aos hábitos culturais de determinada classe e não de qualquer outra» (p. 12). Desta maneira, os exames franceses mais conceituados são *menos provas de conhecimentos do que provas de maneiras ou exercícios de emprego letrado da linguagem*.

Nesta perspectiva, a dissertação continua a ser o meio privilegiado. Coincidindo com a tese sociológica de Durkheim, de Max Weber, de A. Clause, Passeron recorda que o formalismo sempre foi um meio de defesa das classes privilegiadas. É significativo constatar que se dá mais importância a uma prova, a dissertação, que mais radicalmente escapa a toda a classificação objectiva, técnica... Ora, o formalismo consiste precisamente em definir a cultura, não pelo seu conteúdo objectivo, mas pelos imponderáveis que são o aspecto, os cambiantes. Esta «indefinição» permite exercer uma *função de limitação* em proveito dum grupo favorecido.

A neutralidade dos exames científicos não é, aliás, perfeitamente garantida, porque também eles fazem intervir muitas vezes a linguagem, e poucos examinadores ficam indiferentes à elegância da expressão, estranha contudo ao problema técnico em que incide fundamentalmente o exame.

Quanto ao exame oral, reforça-se naturalmente o efeito da boa apresentação da linguagem castigada, da dicção elegante.

A modo de conclusão, podemos observar que a floração dos exames de carácter selectivo parece característica das condições culturais do século XIX, em particular, do desenvolvimento duma burocracia extremamente hierarquizada, ao serviço da

¹ Para um estudo de conjunto, ver *Recherches sur les handicaps socio-culturels de 0 à 7-8 ans*, Bruxelles, Ministère de l'Éducation Nationale, Organisation des Études, 1973.

economia capitalista. É natural, portanto, que a nova forma de civilização para a qual evoluímos seja acompanhada da contestação de um exame concebido para outras condições.

11. Defeito de muitas experiências docimológicas

Há mais de meio século que se acumulam as experiências destinadas a mostrar o desacordo entre examinadores ou a falta de fidelidade de um mesmo examinador.

O problema existe sem dúvida. Mas agrava-se artificialmente, em muitos casos, pelo simples facto de não se convidarem os examinadores a entenderem-se sobre os aspectos a considerar e a importância relativa a atribuir-lhes. Por outras palavras, continua-se a registar parcialmente os resultados duma política anárquica dos exames.

12. Outras críticas

Muitas outras críticas caem ainda sobre os exames. O Relatório da Comissão Consultiva sobre os Exames no Ensino Secundário da Grã-Bretanha, entregue em 1911, apresenta uma lista delas. Aumentamo-la um pouco, com a impressão mesmo assim de não ser exaustiva.

Para o aluno:

- * Consome demasiada energia em reproduzir as ideias dos outros em vez de desenvolver a criatividade.
- * É recompensado por aprendizagens muitas vezes efémeras. (Já se demonstrou que 80% dos conhecimentos, sobretudo factuais, aprendidos para o exame, desapareceram quinze meses depois)
- * Aprende a obediência passiva às palavras de ordem.
- * A aptidão para se exprimir toma mais importância do que o conteúdo. (Mostraremos mais adiante como as desvan-

tagens socio-culturais se acêntuam sobretudo no domínio verbal).

- * Renuncia não raro a exprimir um juízo pessoal para se vergar às ideias do professor que avaliará o exame.
- * A aprendizagem pode ser viciada por um espírito de competição a confinar muitas vezes com o espírito mercenário.
- * Aprende a especular com a sua sorte. No exame tradicional, o pequeno número de perguntas, reflecte idiosincrasias do professor e os alunos não tardam em conhecer os hábitos e as manias dos examinadores.
- * Os exames tradicionais, o longo período de preparação e o período de esgotamento que se segue encurtam consideravelmente o ano escolar efectivo.
- * Os exames tradicionais impedem o trabalho em grupo e exaltam, pelo contrário, o valor da «performance» individual, fonte de egoísmo.
- * A ameaça de fracasso que originam — muitas vezes duplicada por sanções familiares — incita à fraude. O êxito dos batoteiros constitui um mau exemplo para os alunos que são honestos.
- * Os exames dão uma falsa ideia do trabalho adulto, em que quanto mais complexo e difícil é o problema, tanto mais se cercam de conselhos e de obras de referência.

Para o professor:

- * Ensina em função do exame, até mesmo das exigências particulares dos membros dum júri de fora. Veremos que, em boas condições, deve ser o mesmo o espírito do ensino e o do exame. Fazemos alusão aqui a exames estranhos aos objectivos da educação.
- * Sente-se espartilhado na sua metodologia, se o espírito do exame for estranho aos objectivos pedagógicos que lhe parecem essenciais.

Como praticar uma pedagogia da descoberta, da exploração pessoal — processos lentos, mas geradores de aprendizagens profundas — se o exame imposto for de carácter enciclopédico?

- * Escravo do programa, não permite aos alunos avançar segundo o seu próprio andamento e por conseguinte digerir a matéria.
- * Tende a dar demasiada importância às aptidões e conhecimentos úteis ao exame.

Chamou-se até *efeito de repercussão* ao fenómeno que leva os professores a modificar o método e o conteúdo do seu ensino em função da evolução dos exames impostos do exterior.

Podem, assim ser negligenciados — e foram-no muitas vezes — aspectos educativos importantes.

Em resumo, os exames tradicionais apresentam geralmente graves defeitos de construção. A validade é contestável. A avaliação dos trabalhos é, por seu turno, carregada de pesadas deficiências. Além disso, os exames podem ser nocivos à saúde física e mental dos alunos. Acrescentaremos enfim que, em certos casos, é mais em função dos exames, e não o contrário, que os professores fazem os seus cursos...

Mas é tempo de dar a palavra à defesa.

CAPÍTULO 2

DEFESA DA NOTA SUBJECTIVA E DO EXAME

Que se tenha abusado dos exames até ao ponto de viciar a acção educativa da escola, ninguém o poderá contestar. Que muitos concursos não foram mais que sinistras lotarias e ratoeiras parece também fora de dúvida. Que a certos «scores» se tenha atribuído uma significação que eles não tinham é igualmente evidente.

Dá até concluir que exames e concursos devam ser definitivamente proscritos para manter apenas classificações ocasionais e as relações que as sintetizam, parece utópico.

Na primeira parte, vimos que os diferentes modos de avaliação correspondem a funções particulares. Basta retomar a lista para verificar que nem todas podem ser executadas pelo mesmo processo.

Para mais, todos os tipos de avaliação implicam a atribuição dum «score» ou duma nota que se procura que sejam perfeitamente objectivos ou, melhor, perfeitamente controlados. Reconhecemos, pois, aos professores o direito de adaptar muitas das suas avaliações, tanto em função do aluno como da matéria, sob condição de que saibam realmente o que estão a fazer.

Antes de encararmos certas vantagens reconhecidas muitas vezes aos exames, impõe-se uma discussão de carácter negativo.

1. A medida rigorosa é talvez impossível

Procurar a via da perfeita validade dos exames e da fidelidade das avaliações equivale a basearmo-nos em dois postulados recordados por J. Guillaumin¹.

- 1.º «As produções do aluno são pela sua própria natureza mensuráveis, quantificáveis.»
- 2.º «As diferenças que podemos encontrar entre as avaliações praticadas pelos examinadores são susceptíveis de ser reduzidas.»

Se estes dois postulados forem falsos, a docimologia clássica cairá por terra. Ora, muito bons espíritos verificam que, embora as suas grandes linhas estejam definidas há vários decénios, nunca penetraram bem na escola. Este fracasso seria devido em grande parte a uma incompatibilidade entre a natureza psicológica da situação do ensino e o carácter matemático ou físico dos processos de medição propostos.

O debate não é novo e a solução encontra-se no compromisso. Os progressos consideráveis da psicologia e da pedagogia contemporâneas são devidos em larga medida à objectivação da observação, portanto do processo de medir. Contudo, os investigadores das ciências humanas têm hoje uma consciência bastante clara dos seus limites e, em particular, da impossibilidade quase generalizada de utilizar escalas de medição matematicamente perfeitas. Sabem ainda que as mais das vezes medimos apenas comportamentos isolados do todo humano. Como avaliar rigorosamente entidades hipotéticas como o espírito crítico, o sentido da observação, etc., a não ser mediante «performances» particulares supostas representativas do conjunto considerado?

Do mesmo modo nunca se conseguiu, até hoje, medir com precisão o rendimento do ensino, tão numerosos são os pontos de vista a considerar em momentos diferentes, bem

¹ J. GUILLAUMIN, «L'aspect impersonnel de la notation scolaire: de la docimologie à la doxologie», in *Bulletin de la Société A. Binet et T. Simon*, IV, 1968, p. 270.

como não podemos dar-nos conta analiticamente do valor real duma «performance» escolar complexa. Mesmo que, como no cálculo, o resultado possa ser apenas certo ou errado, o raciocínio e o esforço feito podem variar consideravelmente dum caso a outro e, portanto, será difícil de os reduzir a números, objectivamente. Pelo que toca à composição, considerada como obra de arte, escapa à avaliação parcelar¹. Por isso, o direito à subjectividade, à reacção global, conta, como é compreensível, ardentes defensores.

Se a resposta a dar a perguntas de geografia, história, ciências naturais constrange o aluno a exprimir-se, oralmente ou por escrito, deveremos abstrair sempre da elegância da linguagem, da ordenação das ideias, do rigor do raciocínio?

Numa palavra, desde que se deixe de reduzir o êxito ou o fracasso a critérios rígidos e não raro esquemáticos, as produções dos estudantes escapam à quantificação automática, impessoal.

Claro que este argumento não basta para renunciarmos a objectivar uma parte da classificação dos alunos, mas uma parte apenas. A apreciação global do professor, toda a sua agudeza, toda a sua sensibilidade tanto em face da «performance» material como do esforço para ir além e tendo em conta a fraqueza humana, devem manter o seu lugar.

A riqueza do ensino reside antes de tudo na qualidade da relação humana que cria, e a avaliação é um dos aspectos dessa relação. Se ela empobrece ao ponto de não deixar subsistir senão uma comunicação impessoal, então o professor pode ser vantajosamente substituído por uma máquina de ensinar.

¹ Notemos todavia que Ellis Page conseguiu construir um programa de ordenador que permite avaliar automaticamente as composições em língua materna. A correlação com as notas tradicionais é elevada. Seria preciso saber se, em tal caso, o investigador não fez apenas adoptar ao ordenador os passos imperfeitos dos classificadores...

E. PAGE e D. PAULUS, *The Analysis of Essays by Computer*, Washington, U. S. Office of H. E. W., Project 6-1318, 1968.

2. Os professores julgam bem os seus alunos

Se os professores tendem a relativizar os seus juízos em relação ao nível médio do grupo, a sua classificação no interior deste não deixa de ter uma validade elevada.

Bastaria pois aplicar processos de medida que permitissem tornar esses juízos comparáveis entre escolas (veremos como, no capítulo sobre a «moderação») para dispormos de pontos de referência importantes.

A segurança de juízo dos professores explica-se por vários factores:

1. fundam o seu juízo numa observação longa e contínua;
2. consideram um muito maior número de factores (designadamente da personalidade) do que o exame;
3. podem ter em conta particularmente comportamentos excepcionais (em determinada ocasião, certo aluno dá provas duma lucidez, duma originalidade pouco comuns) que não se manifestarão possivelmente no exame.

O valor de previsão dos juízos dos professores (ajustados para os tornar comparáveis) foi posto bem em foco por um estudo feito pela *National Foundation for Educational Research in England and Wales*¹.

Em 1951 e em 1952, cerca de 1200 alunos da região de Wickenham foram examinados na altura da selecção para a entrada no ensino secundário (Grammar Schools). Os resultados foram estudados dois anos e três anos depois.

Provas iniciais:

Testes:

1. Teste de inteligência verbal (V);
2. Teste estandardizado de conhecimentos de língua materna (E1) (Perguntas de escolha múltipla);

¹ Ver F. YATES and PIDGEON, *Admission to Grammar School*, o. c., pp. 57 sg.

3. Teste estandardizado de conhecimentos de aritmética (A) (Cálculos e problemas);
4. Teste de inteligência não verbal (N/V);
5. Teste de aptidão espacial (Esp. 1) (Espaço a duas dimensões);
6. Teste de aptidão espacial (Esp. 2) (Espaço a três dimensões);
7. Teste de conhecimentos de língua materna (E2) (Perguntas mais abertas do que E1).

Juízo formulado por um professor da 6.ª classe da primária

1. Prediz êxito dos estudos na *Grammar School* numa escala de 15 (M);
2. Classifica os alunos segundo a ordem do seu êxito na *Grammar School*. Estes critérios são ajustados pelos investigadores em função dum teste de inteligência verbal (ver descrição do sistema no capítulo da «moderação») (F).

Provas utilizadas para avaliar o êxito no ensino secundário: ordem de mérito fornecida pelo encarregado dos estudos, ajustada em função dos resultados a uma bateria de testes de aptidões e conhecimentos, ministrados imediatamente após a elaboração da classificação (S. M. A.).

O quadro seguinte mostra, para um grupo de 1951, a correlação entre cada prova de previsão e os resultados no decurso dos estudos.

O juízo ajustado dos professores revela-se o melhor meio de previsão. Confirma-se também o valor prognóstico elevado do simples teste verbal.

Convém insistir em que essas observações só são válidas na situação considerada: o êxito na *Grammar Schools*. Não sabemos o que se passaria em outras formas de ensino. Neste caso dispomos, no entanto, duma observação estritamente controlada que nada nos leva a crer que possa ser aplicada a outros casos.

<i>Previsor</i>		<i>Grupo de 1951: dois anos depois</i>	<i>Grupo de 1951: três anos depois</i>
Juízo do professor	(F)	.821	.748
	M	.796	.722
Teste Verbal (V)		.789	.704
E2		.749	.623
A		.734	.659
E1		.729	.622
N/V		.648	.535
Sp.1		.565	.453
Sp.2		.491	.361

3. Validade limitada mas real dos exames tradicionais

~~Fossem quais fossem as imperfeições, os exames tradicionais~~ não impediram a nossa civilização de atingir um nível científico jamais igualado na história da humanidade. A filtragem que operam, revelou-se, portanto, ao menos parcialmente, válida.

E a coisa é evidente. Com a restrição feita no capítulo precedente, podemos afirmar que os exames tradicionais permitiram identificar os mais e os menos dotados, selecção capital na forma de civilização que nos foi dado conhecer nos últimos cem anos.

Que o processo tenha sido grosseiro (por ter ignorado ou desaproveitado talentos, em detrimento sobretudo das classes sociais menos favorecidas) e injusto no que respeita à camada dos alunos médios, mas empurrados brutalmente para o mesmo saco — não invalida em nada o facto de que, numa democracia pouca evoluída, os exames desempenharam bem o seu papel.

4. Criar resistência para a vida

Segundo os behavioristas, não se aprende uma conduta a não ser que efectivamente se realize. Parece pois desejável que, periodicamente, o estudante se veja obrigado a um esforço excepcional, a concentrar a sua energia, a afrontar o ordálio dos exames, com todas as suas imperfeições, até mesmo as suas injustiças. Com efeito, a vida não lhe poupará semelhantes provas e é bom estar preparado para elas.

De mais a mais, a concorrência, a competição são traços fundamentais da nossa civilização. O exame e os concursos existem e é mesmo de desejar que se imponham e acentuam onde há mais candidatos que lugares ou empregos disponíveis; doutra forma o favoritismo sob todas as suas formas e a ineficácia que o acompanha teriam campo livre.

5. Situar-se em relação aos outros

Se uma classificação desfavorável numa prova pode causar um choque, permite também que o estudante se situe em relação aos outros (não em absoluto, mas nas condições do exame!). Aquele que fez o melhor que pode consegue ajustar melhor as suas ambições. O outro terá ocasião de descobrir, antes que seja tarde demais, as consequências das suas fraquezas.

Dum modo geral, o êxito é mais útil que o fracasso, mas uma certa frustração constitui, para alguns, um proveitoso aguilhão.

6. Vasta síntese e integração de conhecimentos

É inegável que os exames, abrangendo vastos domínios de conhecimentos, obrigam o estudante a construir sínteses que o levam a notar, às vezes pela primeira vez, a economia de todo o edifício, as relações entre as partes e, eventualmente, os pontos comuns entre as diferentes disciplinas.

7. O exame externo controla o professor

Um exame regional ou nacional constitui um meio de controlo do trabalho dos professores.

Se já não é admissível sujeitar os professores a programas sobrecarregados, minuciosos até aos pormenores, não é por outro lado menos necessário que o ensino decorra segundo um plano de estudos preciso nos seus objectivos e definido nos conteúdos principais.

Tempos virão em que cada aluno trabalhará segundo um programa individualizado ou semi-individualizado, mas, ainda nesse caso, os fins a atingir deverão ser definidos antes de iniciar a acção pedagógica.

Os pais e as autoridades que organizam e planificam terão sempre o direito de verificar se foi cumprida a missão docente.

A existência de um controlo exercido sobre todos os elementos importantes do plano de estudos parece ainda uma salvaguarda contra a lassidão e o relachamento. Consagrar o tempo necessário à aprendizagem é de louvar; ronçar em ninharias é condenável. Recordo-me de um professor de línguas que, ao fim de dois anos, não tinha passado da introdução do curso... Aparentemente nenhum inspector se apercebeu disso.

Enfim, o exame permite estabelecer, pelo menos em certa medida, se o professor não concentra demasiado os seus esforços em certos alunos em detrimento de outros.

8. O exame externo, «feedback», para o professor

Tanto para a saúde mental como para melhor eficiência, precisamos de ser informados da validade dos comportamentos.

O exame bem concebido permite ao professor emitir um juízo sobre certos aspectos do seu ensino, mediante aprendizagens realizadas pelos alunos.

Seria errado avaliar os professores apenas pelos resultados obtidos a curto prazo. Em todo o caso, só esse género de avaliação é que parece poder fazer-se actualmente com algum rigor científico. Ora, todos os que exercem uma profissão ligada às ciências humanas sentem a dificuldade, embora em graus muito diferentes, proveniente da falta de informação segura sobre a pertinência da sua acção. É essa a preocupação tanto do magistrado como do assistente social, do professor como do sacerdote.

Quando poderá o professor dizer com toda a certeza que cumpriu bem a sua missão?

★
★ ★

Em suma, não somos partidário nem de suprimir completamente os exames, nem de renunciar à classificação subjectiva. O que interessa é adoptar uma solução que aproveite de cada um dos aspectos o que de melhor e mais seguro ele contém. Vamos procurar ver como poderá atingir-se esse objectivo.

TERCEIRA PARTE

CONSTRUÇÃO DO EXAME

INSTITUTO ANTÔNIO AURÉLIO DA COSTA FERREIRA
BIBLIOTECA

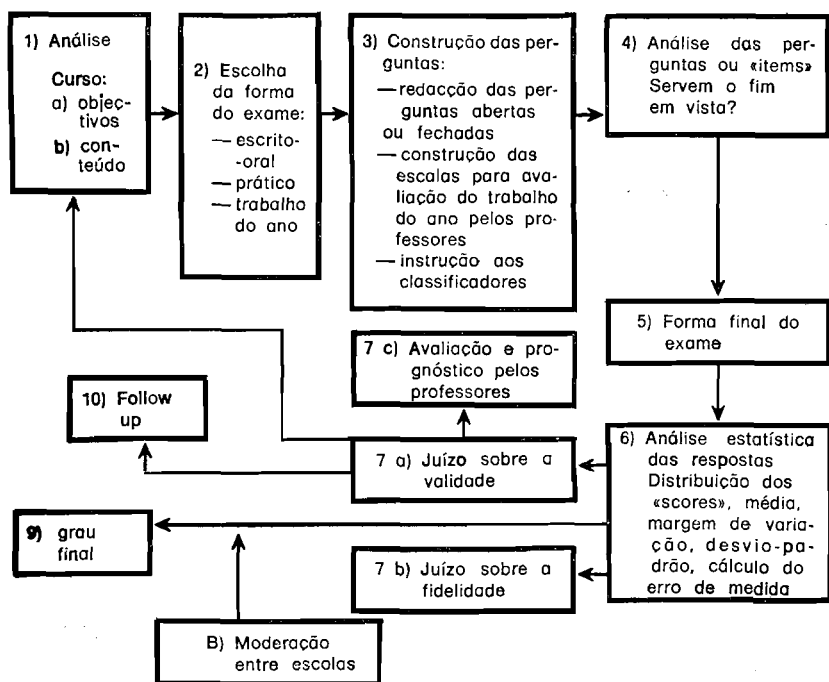


Como os testes estandardizados, os exames preparados pelos professores deveriam construir-se mediante etapas bem definidas. Evidentemente que não é possível, para um simples interrogatório de rotina, realizar os mesmos esforços que para uma prova destinada a centenas de alunos. E, no entanto, as exigências de base continuam as mesmas.

As grandes fases da construção de um exame são as seguintes:

1. definir o objecto e os objectivos;
2. redigir as perguntas;
3. estandardizar a apresentação, a execução e a correcção;
4. aferir;
6. verificar a fidelidade do exame;
6. estabelecer a validade.

O seguinte esquema vai mostrar o caminho a percorrer ¹.



¹ Adaptado de *Examinations Bulletin* n.º 3, 1954, p. 33.

CAPÍTULO 1

O OBJECTO E OS OBJECTIVOS

Ver com clareza a razão de ser do exame que se prepara e definir, com a maior precisão possível, os tipos de aprendizagem a avaliar, é provavelmente a etapa mais decisiva na construção duma prova. Dela depende não só a validade mas ainda a adopção da maneira de avaliar.

I. O objecto

São possíveis três finalidades: o *prognóstico*, o *inventário* e o *diagnóstico*. É raro que uma prova escolar possa incluir-se exclusivamente numa destas três categorias. A maior parte das vezes, os exames pretendem fazer um balanço, o inventário das aquisições, mas ao mesmo tempo desempenham uma função prognóstica a curto e médio prazo (Estará o aluno preparado para iniciar um outro capítulo do curso? Poderá seguir com êxito o curso da classe superior?) e uma função diagnóstica (Onde é que o aluno fraqueja?).

Além disso, importa saber se o exame se insere na acção educativa quotidiana, — caso em que o aluno será comparado a si mesmo e ao grupo a que pertence (a maior parte das vezes, a sua turma) — ou se serve para informar o indivíduo do seu valor relativo na comunidade ou no país, ou para o classificar num concurso, — caso em que deverão ser utilizadas normas adequadas.

A. O prognóstico

A *longo prazo* (pelo menos mais de um ano), o prognóstico formulado a partir dos resultados de um exame escolar é muitas vezes desolador. As mudanças de professor, de método, de espírito de um curso podem provocar uma transformação profunda na atitude do aluno. Além disso, sobretudo nos mais novos, os interesses manifestados mostram importantes flutuações. Devemos ainda ter em conta a evolução psicológica dos alunos. Os professores do ensino secundário, por exemplo, conhecem muito bem a incidência da crise da adolescência no rendimento escolar.

Já abundantemente se demonstrou que, para o prognóstico a longo prazo, os testes de inteligência são indicadores pelo menos tão válidos como os testes de conhecimentos e são muito mais seguros que os resultados escolares¹. Estes testes exploram aptidões que, sobretudo em idades superiores a 8 ou 9 anos, revelam uma grande estabilidade²; dizem respeito ao raciocínio, aos factores verbal, espacial, numérico, ...

Além disso, se as dominantes profundas da personalidade, incluindo as zonas de interesse (não dos interesses particulares, manifestados ocasionalmente) forem identificados, a previsão do êxito escolar e académico pode atingir um nível elevado de segurança.

A *curto prazo*, forma-se de três maneiras o prognóstico escolar:

1) *Testes de maturidade específica (readiness)*

Incidem sobre as formas de raciocínio, sobre as aptidões que intervêm numa aprendizagem determinada — a aprendizagem da leitura, por exemplo. A construção e muitas vezes também

¹ Ver F. HOTYAT (1962), W. McCLELLAND (1945), E. PIEL e D. RUTTER (1951), P. VERNON (1957), F. BACHER (1965), etc.

² Podem ainda obter-se certas modificações. Sabemos hoje que o *Quociente Intelectual* não se fixa duma vez para sempre.

a utilização desses instrumentos complexos dizem respeito ao especialista. Alguns testes que permitem fazer uma primeira sondagem, são todavia concebidos para os professores, como por exemplo o *Teste de Inizan* para a aprendizagem da leitura.

Nos países de língua francesa, seria de desejar que os centros de investigação aumentassem consideravelmente os seus esforços no domínio da maturidade específica. Já mostrámos noutro trabalho como estamos ainda mal equipados para a maior parte dos sectores e a quase todos os níveis¹.

2) *Verificação de conhecimentos-chave* ou *noções críticas* indispensáveis a novas aquisições.

Provas desse género podem ser facilmente organizadas pelos professores, contanto que tenham elaborado sistematicamente a lista dos erros cometidos por ocasião dum ensino anterior².

Também neste caso fichas de investigação, a determinar o requerido anteriormente para abordar os pontos importantes da matéria, prestariam enorme serviço aos professores.

3) *Ensaio*

Para determinar se um aluno está em condições de abordar uma matéria nova, apresentam-se-lhe os primeiros elementos dessa matéria e observam-se sistematicamente as reacções. Uma lição programada de tipo misto (Skinner-Crowder) constitui não raro uma aliança feliz de ensino e de controlo analítico permitindo um diagnóstico a curto prazo.

B. O inventário (provas de rendimento)

Tem por fim, em primeiro lugar, medir as aprendizagens realizadas no decurso de um período mais ou menos longo. Salvo no caso de concursos ou de *surveys* que preparam

¹ Ver G. DE LANDSHEERE, *Les tests de connaissances*, Bruxelles, Editest, 1965.

² Cada noção crítica deve figurar em pelo menos três exercícios. Voltaremos a falar nisso.

reformas, os inventários desempenham também, na escola, uma função prognóstica e diagnóstica.

Com efeito, parece contrário a toda a atitude educativa sã que um professor ao constatar falhas graves não tente identificar a causa para lhe dar remédio e não se inquiete com as dificuldades prováveis que essas falhas indicam para aprendizagens futuras.

Por esse motivo, importa distinguir cuidadosamente, em todo o inventário, as noções críticas indispensáveis para certas aprendizagens ulteriores das noções marginais.

C. O diagnóstico

As provas de diagnóstico tem por fim descobrir e explicar as insuficiências e os hábitos defeituosos em todos os domínios da aprendizagem escolar.

Sem nos aventurarmos na psicologia do fracasso (é, em certos casos, do domínio da psicanálise) notaremos, com D. Durrell¹, que as causas mais frequentes das dificuldades escolares são:

1. a falta de aptidões para realizar uma tarefa;
2. o conhecimento imperfeito dos elementos de base;
3. um ensino mal ajustado ao nível de aptidões da criança e ao seu andamento de aprendizagem;
4. a aquisição de maus hábitos que refreiam o progresso;
5. a inaptidão para transferir, para descobrir o «sistema», para generalizar as técnicas de solução;
6. a falta de vigor, para atacar os problemas, resultante de fracassos repetidos e da falta de interesse.

Se aparecem dificuldades escolares, requerem igualmente a maior atenção tanto o estado físico como a personalidade.

¹ D. DURRELL, *Analysis of Reading Difficulties*, New York, Harcourt, 1955.

Não poderemos esquecer que muitos problemas emocionais, mais do que a causa, são a consequência dos fracassos.

Na prática quotidiana da escola, o diagnóstico preciso e delicado deveria ser uma das preocupações dominantes dos professores e ocupar assim um tempo importante. Ao contrário do que uma espécie de mito da curva de Gauss (ver p. 135) fez crer, o ensino ideal seria aquele que, ao menos para as noções-chave, conduzisse todos os alunos de um grupo a uma «performance» perfeita.

Têm sido publicados alguns bons testes diagnósticos e os professores não podem ignorá-los. Todavia, a maior parte das vezes, poderiam eles próprios construir instrumentos analíticos simples e eficazes, especialmente adaptados às suas necessidades.

É fácil imaginar a utilidade de um quadro de conjunto em que, para a maior parte dos alunos duma turma, se põe em evidência com + ou com — se a resposta foi ou não correcta para três exercícios que incidem sobre um dos pontos da matéria¹. Basta colorir a vermelho as quadrículas que contêm dois ou três sinais menos para obter uma primeira impressão da situação². Há anos, com a colaboração dos alunos, organizámos e pusémos em prática um quadro desse género para a pronúncia duma língua estrangeira. Os resultados foram positivos.

Para que a prova possa desempenhar um papel diagnóstico, é preciso evidentemente que os exercícios incidam sobre um aspecto muito preciso da matéria e que sejam concebidos, se possível, por forma a não suscitarem mais que uma dificuldade de cada vez.

Como a elaboração de semelhantes controlos se torna facilmente pesada, é muitas vezes aconselhável a colaboração de vários professores.

¹ Numa prova diagnóstica, verifica-se três vezes pelo menos o conhecimento da mesma noção, em pontos diferentes do teste, a fim de reduzir seriamente o papel eventualmente desempenhado por uma distração momentânea, por uma resposta dada ao acaso, etc.

² Para um exemplo de aplicação deste sistema, ver: BONGRAIN, e al., *Epreuves analytiques d'arithmétique* (fim da primária), Institut Supérieur de Pédagogie, 1961.

II. Os objectivos

«É curioso verificar que, ao pedirmos às pessoas que têm a seu cargo a educação das crianças para definirem os fins que se propõem, mergulhamo-las não raro na perplexidade».

P. Osterrieth, *Faire des adultes*, p. 9

Afirmar que os professores devem não só *instruir* mas também *educar* tornou-se já um truísmo. Importa no entanto não se ficar em votos piedosos, em preocupações vagas. Ensinamos evidentemente o que somos: a nossa personalidade, a nossa maneira de pensar e de agir influenciam directamente, e quase sem darmos por ela, os nossos alunos.

Mas, se quisermos sistematizar a nossa acção, torna-se indispensável uma definição precisa dos fins a atingir.

Educar é pôr em jogo os meios adequados ao desenvolvimento da inteligência e da personalidade no sentido que o meio cultural indica, sem esquecer que a saúde física exige também atenção do educador.

Instruir significa tornar alguém possuidor de conhecimentos novos.

Os *objectivos gerais* dizem respeito à educação, os *objectivos especiais* são dos domínios da instrução. Mas ambos são inseparáveis ou pelo menos deviam sê-lo.

E é *antes de começar o ano* escolar que o professor deve interrogar-se sobre os objectivos a atingir (do contrário, anda às cegas) e *definir também a matéria dos exames*.

É certo que os planos de estudos e os programas escolares **oficiais** dão já uma parte das respostas. Mas só depois de um estudo minucioso e duma longa meditação sobre esses documentos de base é que essas respostas irão aparecer. Aliás não é raro, sobretudo no ensino secundário, que cada secção seja **objecto** de um programa separado e que só sejam explicita-

mente considerados os objectivos privilegiados. Pertence portanto ao professor acrescentar os objectivos mais gerais que irá procurar atingir em comum com os outros colegas das diferentes disciplinas. Se o professor não tiver uma consciência perfeita dos objectivos e não os tornar seus, tem poucas probabilidades de os atingir.

Estabelecer um quadro dos objectivos a atingir ajuda muito o professor na altura em que prepara a matéria do ano. Esse (ver por exemplo a p. 75) terá tantas colunas quantas os objectivos gerais. Para cada capítulo ou etapa do programa, indicam-se os objectivos que se deseja atingir e a importância relativa que se lhes irá atribuir (escala de 3 a 5 graus).

A apreciação da importância relativa dos diferentes pontos do programa escapa à quantificação rigorosa. Tratar-se-á duma matéria de base que se deve necessariamente conhecer para poder abordar as outras? Ou serão antes conhecimentos ou capacidades indispensáveis para o exercício duma profissão ou para a inserção na vida?

Por subjectiva que a avaliação da importância possa às vezes ser, não há dúvida que o esforço de reflexão que exige conduz quase sempre a uma útil clarificação.

Frequentemente, os educadores fazem a si mesmo uma pergunta mais grave: «Os objectivos que pretendem alcançar não vão trair os resultados fundamentais, os fins da educação?». Toda a contradição entre a filosofia e a educação, particularmente os valores adoptados, e a acção pedagógica deve ser evitada.

A. OS OBJECTIVOS GERAIS

No fundo, o objectivo do ensino é único: formar o adulto. P. A. Osterrieth consagrou a este tema páginas penetrantes¹.

¹ P. A. OSTERRIETH, *Faire des adultes*, Bruxelles, Dessart, 1964. Ver particularmente o capítulo I: «O fim da Educação».

Mais particularmente, os professores devem, por um lado, ao menos não prejudicar a saúde física e mental dos alunos e, quanto possível, ajudá-los a desenvolver-se, e, por outro lado, procurar obter sistematicamente objectivos cognitivos e afectivos que vamos tentar definir.

Separar os domínios cognitivo (pensar), afectivo (estar bem disposto ou irritado; amar ou desprezar) e conativo (querer, desejar), é muito teórico. Poder-se-á pensar sem intervenção do sentimento, agir sem pensar? Todos sabemos que é o organismo total que responde a um estímulo e, para cada comportamento, uma análise subtil deveria revelar a importância relativa das componentes cognitiva, afectiva e psicomotriz.

A nossa intenção é bem mais modesta. Andamos actualmente à procura de pontos de referência que servirão para escalonar o ensino e os exames. É por isso que vamos separar arbitrariamente o domínio cognitivo do afectivo e dar a esta última palavra uma acepção mais vasta.

1. Objectivos cognitivos

É a dois investigadores americanos, B. Bloom¹ e J. P. Guilford², que nós devemos classificações hierarquisadas dos objectivos cognitivos que, não obstante certas falhas³, se revelam instrumentos úteis.

a) Taxonomia de Bloom

Em primeiro lugar, as grandes articulações. A página 77 propõe-se um exemplo de aplicação prática.

1. Saber de cor
2. Compreender
3. Aplicar

¹ B. BLOOM e al., *Taxonomie des objectifs pédagogiques. I. Domaine cognitif*, traduzido por M. Lavallée, Montréal, Education Nouvelle, 1969.

² J.-P. GUILFORD, *Modèle tridimensionnel de l'intellect*. Foram publicadas cerca de quarenta monografias sobre o assunto por Guilford e a sua equipa (University of South California, Los Angeles).

³ Ver particularmente a crítica de J. CARROLL.

4. Analisar
5. Sintetizar
6. Avaliar.

Cada um destes escalões deve ser aceite de acordo com o sentido dado pelos autores da taxonomia¹. Além disso, mesmo que o saber de cor e o avaliar surjam como comportamentos cognitivos menos nobres ou mais nobres, não quer de modo algum significar que a memorização deva desaparecer do ensino. Teoricamente pelo menos, é preciso ter ultrapassado escalão por escalão para ter acesso a um nível superior. A taxonomia foi criada para *ajudar o educador a não esquecer* certas etapas e para o incitar a elevar gradualmente o seu nível de ensino.

1. SABER DE COR

- 1.1. Dados particulares: designações, factos, datas, símbolos.
- 1.2. Maneiras de tratar dados particulares (sem os aplicar): convenções, classificações, critérios, métodos.
- 1.3. Dados universais: princípios, leis, teorias...

A diferença entre as três subcategorias é mais quantitativa que qualitativa. Saber de cor, sem mais, as datas do nascimento dos reis de França ou determinada exposição da filosofia de Kant exige sobretudo esforços de memorização diferentes. Em ambos os casos, o nível do pensamento permanece muito baixo. Ora quantas perguntas de exame, desde a primária ao ensino superior, se situam nesse nível (Que sabe acerca de...? Quais são as cláusulas de...? Como se prepara o...?).

2. COMPREENDER

Trata-se aqui da compreensão ao mais baixo nível. Mediante uma formulação nova do dado, o indivíduo mostra que ultra-

¹ Aplicar, por exemplo, implica uma análise mais ou menos delicada. E por isso não seria de colocar a análise antes da aplicação? Bloom e os seus colaboradores reduziram o sentido de *aplicação* para escapar à dificuldade. Indicam, aliás, em diversos passos da obra, a interferências entre alguns escalões e analisam com muita agudeza as dificuldades de aplicação da *Taxonomia*. O nosso objectivo não é reproduzir todos os matizes, mas, chamar a atenção para o instrumento.

passou o psitacismo, que a mensagem tem, para ele, uma significação. Em todo o caso, não se trata ainda de aplicar, portanto de notar as relações do 'dado com outros materiais, com outras situações.

Bloom distingue dois escalões:

2.1. Traduzir, transpor

Conserva-se o conteúdo da comunicação sem que seja modificada a ordem, mas modifica-se a forma.

Exemplos: parafrasear uma narração, uma oração: «Uma *taxonomia* é uma *classificação*»; exprimir verbalmente símbolos matemáticos:

$A > B$ significa que A é maior que B .

2.2. Interpretar

Explicar ou resumir uma comunicação. A interpretação implica um novo arranjo, uma nova visão do material. Supõe, portanto, a capacidade de reconhecer e de entender as ideias fundamentais duma comunicação e de compreender as relações que entre elas existem. «Sob este ponto de vista, a interpretação torna-se sinónimo de análise e possui certas características da avaliação»¹.

Exemplo: Interpretar dados representados sob a forma de quadros ou gráficos tirando deduções que tenham em conta relações entre os dados ou a sua significação de conjunto.

Ir além dos dados ou das instruções recebidas: extensão das tendências, generalização.

2.3. Extrapolar

3. APLICAR

A aplicação supõe que o sujeito distinga os traços comuns a duas situações, a dois problemas. Originou-se portanto uma abstracção.

¹ B. BLOOM e al., o. c., p. 104.

Exemplo: utilização de processos experimentais para resolver problemas de trabalhos caseiros.

4. ANALISAR

4.1. Busca de elementos.

Exemplo: Distinguir, numa comunicação, os factos das hipóteses.

4.2. Procurar as relações.

Exemplo: Serão lógicas as hipóteses em relação aos informações de que se dispõe?

4.3. Procurar princípios de organização.

Exemplo: Identificar as técnicas de propaganda utilizadas em folhetos.

5. SINTETISAR

5.1. Produzir uma obra pessoal.

Exemplo: Narração apaixonante duma experiência vivida.

5.2. Elaborar um plano de acção correspondente a exigências fixadas.

5.3. Derivar um conjunto de relações abstractas. Induzir uma regra.

6. AVALIAR

Juízos qualitativos e quantitativos estabelecem em que medida o material e os métodos correspondem aos critérios (internos ou externos).

Exemplos: Descobrir os sofismas numa discussão. Apreciar um trabalho por comparação com um modelo.

b) O modelo de J. P. Guilford

Menos utilizado até hoje que a taxonomia de Bloom para a construção dos exames, o modelo de J. P. Guilford oferece talvez maiores possibilidades ainda, em virtude do seu rigor.

Vejamos em primeiro lugar como J. P. Guilford e R. Marrifels definem as três dimensões do intelecto e as suas componentes¹.

AS OPERAÇÕES

São as actividades ou processos intelectuais mais importantes. É o que o organismo faz a partir da matéria prima informacional, a partir do que ele discrimina.

1. Cognição

Consciência, apreensão, descoberta ou redescoberta, reconhecimento, compreensão de informações sob diversas formas.

2. Memória.

Retenção de informações.

3. Produção convergente.

O gerar de informações únicas, convencionalmente aceites, a partir de um dado. São respeitados o uso, o costume e a regra.

4. Produção divergente

O gerar de informações variadas a partir de um mesmo dado. Originalidade, criatividade,

¹ Ver as *Reports from the Psychological Laboratory of the University of the Southern California*, publicados a partir de Junho de 1950.

5. Avaliação

Tomada de decisões ou formulação de juízos respeitantes à exactidão, a adequação, ao que é desejável, ... em conformidade com critérios, ideais, objectivos adoptados.

OS CONTEÚDOS

1. Figurativos

Informação na sua forma concreta, apercebida ou recordada em imagens.

Um mínimo de organização, de estruturação é indispensável. Inteligência prática.

2. Simbólicos

Informações sob forma de sinais desprovidos de significação em si mesmos: letras, números, notas de música.

Inteligência teórica.

3. Semânticos

Informações sob forma de significações ligadas a palavras. Inteligência verbal.

4. De comportamento

Informações essencialmente não verbais que intervêm nas interacções humanas em que a percepção de atitudes, de necessidades, de desejos, de intenções, de pensamentos de outrem ou de si mesmo desempenham um papel.

Inteligência social.

OS PRODUTOS

São os resultados do tratamento das informações pelo organismo.

1. Unidades

Porções de informação relativamente isoladas ou circunscritas.

2. Classes

Unidades agrupadas em virtude das propriedades comuns.

3. Relações

Conexões reconhecidas entre unidades.

4. Sistemas

Agrupamento de unidades organizadas ou estruturadas; complexos de partes em interrelação ou interacção.

5. Transformações

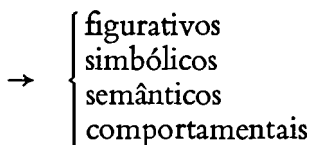
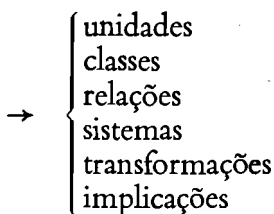
Mudanças introduzidas nas informações ou na sua utilização.

6. Implicações

Extrapolação de informações: predição, consequências, antecedentes.

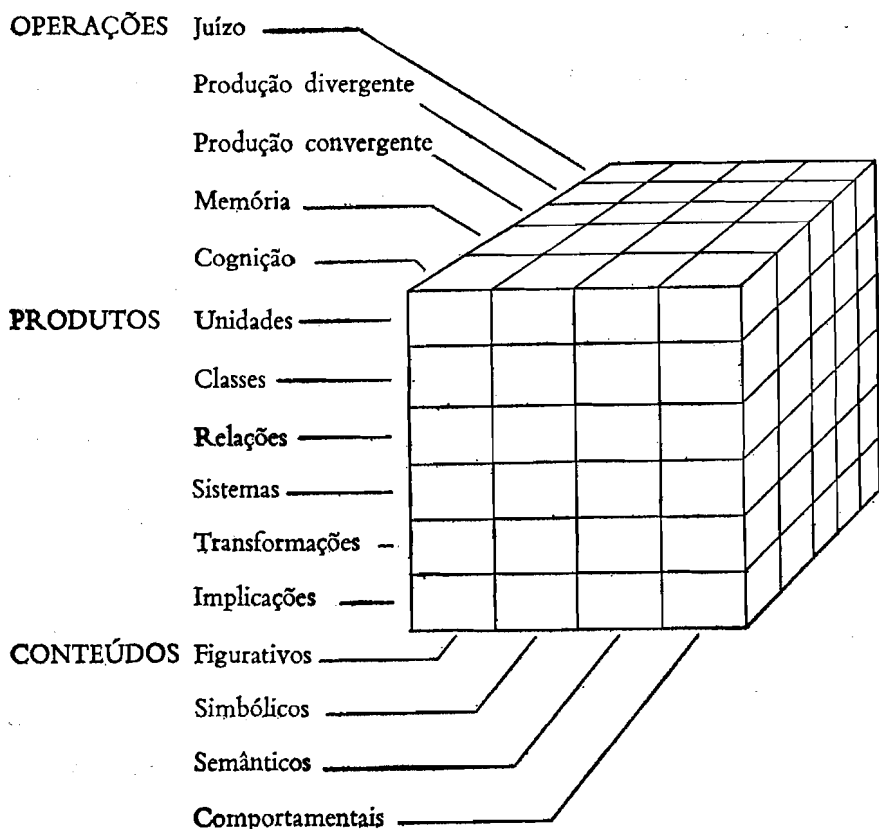
Cada uma das componentes das três dimensões combina-se com todas as outras. E temos:

Memória das



ou seja 24 combinações.

Como se distinguem cinco tipos de operações, existem ao todo, portanto, cento e vinte combinações no modelo.



A experiência mostra que é difícil traduzir todas as combinações em termos utilizáveis no ensino.

Num primeiro momento, irão reter a atenção sobretudo os quatro tipos de operações. Do ponto de vista ideal, e desde o início da escolaridade, o professor deveria esforçar-se por introduzi-los em todas as actividades.

1. A reprodução de cor indica a fixação da noção.
2. A produção convergente indica a capacidade de aplicar as aquisições em conformidade com as regras, os costumes.

3. A produção divergente indica a capacidade de descobrir soluções ou usos novos.
4. Enfim, a avaliação indica a capacidade de criticar, de emitir um juízo em função de critérios bem definidos e de agir em consequência disso.

Muitas vezes a actividade escolar e os exames que a sancionam limitam-se às duas primeiras operações.

Esforçar-se por manter um equilíbrio entre as quatro categorias de conteúdos é tão importante no ensino geral como no técnico ou artístico.

2. Os objectivos afectivos

Esquematisando a análise penetrante de Osterrieth ¹, dizemos que um indivíduo atinge plenamente a condição de adulto:

- 1.º a) se o seu comportamento encontrou a lógica, a coerência e escapa à versatilidade;
b) se, além disso, adquiriu uma sã tolerância perante a mudança, a contradição, o fracasso;
- 2.º a) se conquistou a independência, a autonomia intelectual e afectiva;
b) se no entretanto é capaz de se doar, de permanecer fiel aos seus compromissos e sentimentos.

A educação é uma longa caminhada para este último equilíbrio.

Proporemos mais adiante uma adaptação e uma interpretação da taxonomia dos objectivos de D. Kratochvíl e B. Bloom ². Obtêm-se de novo cinco escalões, cinco etapas que levam do comportamento mais passivo ao mais activo.

¹ P. OSTERRIETH, *Faire des adultes*, Bruxelles, Dessart, 1964.

² D. R. KRATOWOHL, B. S. BLOOM, B. MASIA, *Taxonomie des objectifs pédagogiques, Domaine affectif*, Montréal, Education nouvelle, 1970.

a) **O indivíduo responde a um estímulo exterior**

1. *É simplesmente receptivo*

É uma espécie de estado receptivo amorfo em que o sujeito se apercebe da beleza e da fealdade, dos sentimentos diversos, sem reagir, quase como um espelho que não reflectisse a imagem.

Este comportamento é, aliás, difícil de distinguir da simples cognição que precede a memorização. Só se pode observar um certo despertar da atenção. Exemplo: ouve música, ouve falar dos outros.

2. *Recebe e reage*

O indivíduo reage nitidamente, quer obedecendo, quer manifestando prazer, ou pela palavra ou pelo gesto ou pela atitude. Neste estágio, não se observa ainda rejeição explícita que testemunharia uma escolha deliberada.

Para o professor de literatura, é o momento em que os alunos não têm ainda o gosto bastante formado para fazerem uma escolha pessoal, em que a sua sensibilidade não está ainda suficientemente afinada para lhes permitir partirem sozinhos à descoberta, mas em que, postos em presença de obras de valor, começam a sentir-lhes a grandeza.

3. *Recebe e reage, aceitando ou recusando*

Agora o indivíduo sabe o que quer e do que gosta. Basta pô-lo em presença das pessoas ou das coisas. Empenha-se.

b) **O indivíduo toma a iniciativa**

4. *Tenta espontaneamente compreender, julgar, sentir*

O indivíduo sente bastante interesse, curiosidade por se instruir sem para isso ser solicitado, bastante sensibilidade para tomar uma iniciativa sentimental ou então descobrir suficiente-

mente o sentido dos valores para escolher uma filosofia ou uma religião.

5. *Age segundo as suas opções*

É o estágio psicologicamente adulto, como o definiu P. Osterrieth.

O indivíduo, por exemplo, vive em função das suas opções morais, sentimentos, estéticas, mas também é capaz de mudar de conduta à luz de provas, de argumentos convenientes.

Esta última etapa da ascensão afectiva corresponde à avaliação no domínio cognitivo.

B. OS OBJECTIVOS ESPECIAIS

Teoricamente, todas as matérias, todos os pontos dos programas escolares dão aso a que nos aproximemos dos objectivos gerais e verifiquemos se foram atingidos.

No entanto, três observações se impõem.

- 1.º A prática mostra que uma mesma prova não pode comportar senão um número limitado de objectivos.
- 2.º Certas disciplinas prestam-se melhor que outras a consecução de determinados objectivos.
- 3.º Continua a pôr-se com acuidade o grave problema do *transfert* das aprendizagens. Cultivar, por exemplo, a divergência em face das actividades artísticas em nada garante que esse traço se venha a manifestar nos domínios científicos ou na vida prática.

Seja como for, o primeiro passo com vista à identificação dos objectivos especiais permanece sempre fundamentalmente o mesmo e consiste na elaboração de um quadro de dupla entrada. No alto, indicam-se os objectivos gerais; ao lado, as matérias do curso. Cada intersecção de coluna e de ordem indica, em princípio, um objectivo especial.

Exemplo: QUÍMICA ¹

	Conhecimento dos factos, métodos e técnicas	Aplicação	Avaliação
Equilíbrios iónicos ¹			
1. Generalidades			
1. Grau de ionização ou fracção ionizada			
2. Medidas do grau de ionização			
3. Aplicação da lei de acção das massas ou equilíbrios iónicos .			
4. Lei da diluição de Ostwald .			
2. Produto de solubilidade			
1. Definição			
2. Cálculo do produto de solubilidade em função da solubilidade			
3. Diminuição da solubilidade de um electrólito pouco solúvel .			
4. Precipitação de um electrólito por adjunção, à solução saturada, de um electrólito com um ião comum			
5. Dissolução dos precipitados .			
<i>Apêndice:</i> Aplicação do produto de solubilidade à precipitação dos sulfuretos			
3. Produto iónico da água e pH . .			
1. Produto iónico da água . . .			
2. O pH			
<i>a</i> A notação pH			
<i>b</i> pH da água pura			
<i>c</i> A escala dos pH			
<i>d</i> Determinação do pH e os indicadores corados			
<i>f</i> Métodos experimentais de determinação do pH			
<i>g</i> Distinção entre electrólitos (ácidos e bases) fortes e fracos, por determinação do pH de soluções diluídas de concentrações conhecidas . .			

¹ Extraído de BERGER e DIGHAYE, *Chimie IV*, Liège, Sciences et Lettres, 1967.

	<i>Conhecimento dos métodos e técnicas</i>	<i>Aplicação</i>	<i>Avaliação</i>
<p><i>h</i> Cálculo do pH das soluções de ácidos e de bases a partir das suas concentrações molares e, eventualmente, das suas constantes de ionização . .</p> <p>4. Soluções tamponadas</p> <p>1. Introdução experimental</p> <p>2. Cálculo do pH duma mistura tampão</p> <p>5. Hidrólise</p> <p>1. Generalidades</p> <p>2. Caso de sais que derivam dum ácido forte e duma base forte .</p> <p>3. Caso de sais que derivam dum ácido forte e duma base fraca</p> <p>4. Caso de sais que derivam dum ácido fraco e duma base forte</p> <p>5. Caso de sais que derivam dum ácido fraco e duma base fraca</p> <p>6. Factores que influenciam a hidrólise</p> <p>6. Neutralização mútua dos ácidos e das bases</p> <p>1. Calor de neutralização</p> <p>2. Curvas de neutralização</p> <p> <i>a</i> Ácido forte-base forte</p> <p> <i>b</i> Ácido fraco-base forte</p> <p> <i>c</i> Base fraca-ácido forte</p> <p>7. Os métodos de análise quantitativa por via química</p> <p>1. Métodos gravimétricos</p> <p>2. Métodos volumétricos</p> <p> Ácidos-bases</p> <p> Escolha dos indicadores</p> <p> Titulação redox</p>			

A matéria de química que figura no quadro anterior constitui um dos quatro capítulos de um curso destinado ao ano superior do ensino secundário. Entre muitos outros objectivos gerais possíveis, apenas indicámos três.

Um simples golpe de vista sobre o conjunto revela que, mesmo nessas condições extremamente simplificadas, o professor se encontra perante uma escolha difícil. Sobre que pontos irá fazer incidir o exame para obter uma amostragem suficiente da matéria? Que importância relativa irá reservar às matérias consideradas e aos diversos objectivos especiais que lhes correspondem? Em que consistirá exactamente a avaliação num curso como este?

Como vemos, o problema está longe de ser simples. Ultrapassa, quase sempre, a competência de um homem só e leva a perguntas que permanecem sem resposta. Seria desejável que se criassem comissões para definição de objectivos em que docentes, psicólogos e especialistas da investigação pedagógica, unissem os seus esforços.

Vamos ver agora uma série de *items* de ciências que ilustram os principais níveis da hierarquia de B. Bloom.

TAXONOMIA DE BLOOM

Ilustração: Ciências naturais¹

1. *Conhecimento*

(a) Conhecimento de dados específicos

Os glucidos e os lípidos são utilizados pelo organismo:

- A. como alimentos energéticos;
- B. Para formar novo citoplasma;
- C. como fonte de sais minerais;
- D. como alimentos plásticos;
- E. como fonte de vitaminas.

¹ Adaptado de D. G. LEWIS, «Objectives in the Teaching of Sciences», in *Educational Research*, June 1965, pp. 186-199.

- (b) Conhecimento da maneira de tratar informações específicas

Equilibre a equação: $\text{H}_2\text{S} + \text{O}_2 \rightarrow \text{H}_2\text{O} + \text{SO}_2$
e conte o número de moléculas para o conjunto.

- A. 4
- B. 18
- C. 5
- D. 9

E. outro número diferente dos anteriores.

- (c) Conhecimento dos princípios gerais

Uma máquina simples não pode:

- A. mudar a direcção duma força aplicada;
- B. substituir uma força pequena por uma grande;
- C. produzir um resultado de 100 entre as forças postas em jogo;
- D. vencer uma resistência maior que a força aplicada;
- E. aumentar a energia do sistema.

2. *Compreensão* (o aluno conhece e sabe utilizar a matéria que lhe foi comunicada sem poder necessariamente estabelecer relações com outras matérias).

Se, em tempo muito frio, uma criança põe a mão num recipiente de aço, pode acontecer que a mão adira, o que não aconteceria se o recipiente fosse de madeira.

É porque o aço:

- A. tem um ponto de fusão mais elevado que a madeira;
- B. irradia mais o calor que a madeira;
- C. tem uma massa específica mais elevada que a madeira;
- D. é melhor condutor do calor que a madeira;
- E. possui propriedades magnéticas, ao passo que a madeira não.

3. *Aplicação* (abstrair duma situação particular e aplicar a abstracção duma outra maneira).

Quando o magnésio se queima para formar óxido de magnésio, MgO , combinam-se dois gramas de oxigénio com 3 gramas de magnésio. Se queirmos 6 gramas de magnésio num recipiente fechado contendo 3 g de oxigénio:

- A. não haverá reacção;
- B. formar-se-á Mg_2O ;
- C. formar-se-á MgO_2 ;
- D. ficará uma certa quantidade de magnésio;
- E. o óxido de magnésio conterá uma menor percentagem de oxigénio.

4. *Análise*

Se dois objectos, X e Y têm o mesmo momento (momentum), a energia cinética de Y, não pode ser superior à de X a não ser que:

- A. Y seja mais pesado que X;
- B. Y se desloque mais depressa que X;
- C. Y tenha o mesmo peso que X;
- D. Y se desloque mais lentamente que X;
- E. Y esteja situado a uma distância maior do eixo que X.

5. *Síntese* (combinar elementos para formar um todo estruturado) ¹

Um certo número de problemas que retiveram a atenção dos homens de ciência podem classificar-se na rubrica «estrutura da matéria». Neste domínio, as primeiras investigações incidiram sobre a estrutura descontínua da matéria, hipótese emitida por Dalton em 1805.

¹ Este item deve-se a H. Guillaume.

Um outro grupo de investigações considera a teoria cinética dos gases. Aparece já no séc. XVII, retomada depois de forma mais completa por Maxwell (1859) e Boltzmann (1896). Mostre como a teoria cinética dos gases constituiu um argumento favorável à hipótese da representação atomística da matéria.

A sua discussão deve compreender:

1. uma exposição das hipóteses fundamentais da teoria;
2. algumas deduções importantes dessa teoria, aplicada ao caso dos gases perfeitos;
3. a interpretação das leis experimentais de Gay-Lussac e da hipótese de Avogadro.
6. *Avaliação* (juízos de valor sobre os argumentos que servem a uma demonstração ou sobre critérios).

A teoria atômica de Dalton é:

- A. um bom exemplo de verdade absoluta;
- B. um bom exemplo duma concepção estática da ciência;
- C. exacta e permitiu outras descobertas científicas;
- E. nem correcta nem incorrecta e ninguém se tem ocupado dela.

Finalmente, alguns exemplos de perguntas, tendo em conta os três aspectos característicos do modelo de J. P. Guilford:

<i>Ramo</i>	<i>Produção convergente</i>	<i>Produção divergente</i>	<i>Avaliação</i>
Ciências	Explique por que não pode haver vida em Mercúrio	Em que é que a vida em Marte poderia diferir da nossa?	Parece-lhe que existe vida em Marte?
Geografia	Em que é que o Estreito de Bering influenciou o povoamento da América do Norte?	O que aconteceria se o Estreito de Bering não tivesse existido?	Actualmente qual é, na sua opinião, o principal papel do Estreito de Bering?
História	Explique a importância do descobrimento da América por Cristóvão Colombo para a vida económica da Europa.	Que teria acontecido se Cristóvão Colombo tivesse descoberto o Caminho das Índias em vez da América?	Quais são, na sua opinião, as duas consequências mais importantes da viagem de Colombo?
Língua materna	Explique porque é que a novela se espalhou muito mais rapidamente nos Estados Unidos do que na Europa.	Aqui tem o princípio duma notícia. Imagine tantos desfechos quantos puder.	O que é que mais importa num romance: os caracteres ou a intriga?

¹ Segundo J. R. VERDUIN JR., *Conceptual Models in Teacher Education*, Washington, A. A. C. T. C., 1967, p. 93.

C. OS OBJECTIVOS OPERACIONAIS¹

Como vamos ver, o recurso às taxonomias não exclui a subjectividade. A prova é que, para cada intersecção do quadro matérias-objectivos (p. 86), diversos examinadores trabalhando isoladamente formularão perguntas por vezes muito diferentes. Além disso, é preciso provar que uma pergunta determinada põe em acção comportamentos comandados pelo objectivo.

O problema fica já bastante clarificado se os objectivos forem directamente expressos em termos de comportamentos observáveis.

Comparar:

— O aluno deve conhecer o nome dos reis da Bélgica que se sucederam desde 1830 (objectivo não comportamental).

— O aluno nomeará correctamente os reis belgas... (objectivo comportamental).

— O aluno irá adquirir um conhecimento activo do gravador (objectivo não comportamental).

Dispondo de um gravador e de uma fita registadora virgem, o aluno registará a sua própria voz (objectivo comportamental).

Em ambos os casos o verbo «conhecer» pode ser diversamente interpretado. Ao contrário, «nomear» ou «registar» são desprovidos de ambiguidade.

Notaremos ainda que os objectivos formulados dessa maneira contêm em si mesmos a pergunta que permitirá verificar se são atingidos.

Sob o impulso particular de W. Popham² e de um modo geral dos especialistas da tecnologia do ensino, os esforços

¹ Segundo H. Piéron, a definição operacional dum comportamento, dum factor, etc., é «o enunciado do procedimento que permitem medi-lo, produzi-lo, ou simplesmente reconhecê-lo entre outros». (Vocabulaire de la psychologie, Paris, P. U. F., 1968).

² W. POPHAM, *Objectives and Instruction*, Chicago, Rand Mc Nally, 1967.

de clarificação ao nível dos micro-objectivos, foram consideráveis nestes últimos tempos.

Os micro-objectivos não bastam no entanto para constituir uma unidade de ensino. Tão estreitamente limitada, tornar-se-ia artificial. Ao contrário, verificar se uma bateria de micro-objectivos relativos a um mesmo objecto são atingidos dá-nos um processo diagnóstico de primeira apanha.

Dois exemplos

1. *Biologia — ensino primário*

Lista publicada pela «Bolsa de objectivos» animada por Popham (extracto)¹.

Categorias principais

- I. O organismo individual — 7 princípios — 46 objectivos específicos.
- II. A população — 5 princípios — 25 objectivos específicos.
- III. O meio ambiente — 2 princípios — 10 objectivos específicos.
- IV. A comunidade — 3 princípios — 11 objectivos específicos.
- V. O ecossistema — 4 princípios — 16 objectivos específicos.

Divisão da categoria I: o organismo individual

Princípio A: os objectos são classificáveis em vivos e não vivos.

¹ A «Bolsa de objectivos» (Instructional Objectives Exchange) foi criada em 1968, pela U. C. L. A. O seu objectivo é:

- servir de clearinghouse que permita a todas as escolas do país trocar os objectivos de ensino; deste modo, em vez de se dispersarem, conjugam-se os esforços dos educadores;
- reunir e construir técnicas de medida que permitam determinar se os objectivos são atingidos;
- formular de maneira apropriada objectivos de ensino em sectores importantes que ainda os não possuem; trata-se de preencher lacunas existentes.

Objectivo específico A₁

Numa série de desenhos, o aluno deverá saber distinguir os objectos vivos.

Objectivo específico A₂:

Apresenta-se o desenho duma semente não germinada. O aluno deverá saber explicar por escrito ou oralmente como se pode saber se a semente está morta ou viva.

Objectivo específico A₃:

O aluno deverá saber distinguir as plantas e animais numa série de desenhos.

Objectivo específico A₄:

Numa série de perfis de animais, o aluno deverá saber indicar os animais que possuem vértebras; deverá saber desenhar a coluna vertebral mediante uma linha sobre o perfil.

Etc.

Princípio B:

Para obter energia e preencher funções necessárias à vida, os organismos buscam e utilizam uma variedade de substâncias nutritivas. A energia extraída dessas substâncias é armazenada no organismo, em seguida é utilizada para o crescimento, movimento e reprodução.

Objectivo específico B :

Saber reconhecer num desenho as fontes de energia que convêm aos animais que nele se encontram.

Etc.

2. *Ensino da leitura e da escrita*¹

Vamos apresentar uma das vinte e quatro baterias de objectivos relativos ao ensino elementar da leitura.

¹ W. POPHAM, *Language Arts: Decoding Skills*, Los Angeles, Instrucional Objectives Exchange, 1972.

OBJECTIVO D

O aluno terá de discriminar as letras entre si

- D₁ Partindo de um conjunto de letras que diferem no tamanho e no desenho, o aluno terá de identificar as que se correspondem.

Ex.:

A B C **A** Δ B

- D₂ Partindo de diferentes letras que têm um elemento comum, o aluno terá de identificar o que as diferencia.

Ex.: d/b ; g/j ; m/n ...

- D₃ Partindo dum algarismo ou duma letra modelos, o aluno terá de identificar numa série de algarismos ou de letras o elemento idêntico ao modelo dado.

Ex.: b: d b p q.

- D₄ Partindo duma letra ou grupo de letras escritas, o aluno terá de sublinhar, numa frase dada, todas as letras idênticas à letra ou às letras propostas.

Ex.: Ca: Carlos cabe na cama.

- D₅ A partir duma lista de palavras que têm diversas letras iniciais ou finais, o aluno deverá classificar as palavras com inicial ou final idênticas.

- D₆ A partir duma sequência de letras seguida duma série de sequências de letras, o aluno irá referenciar a sequência idêntica à sequência dada como modelo.

Ex.: par: por pra rap par.

CAPÍTULO 2

REDACÇÃO DAS PERGUNTAS

I. Observações gerais

Para ser equitativo e válido, um exame deve quase sempre conter um grande número de perguntas. Se o domínio a ter em conta for extenso, não parece haver dúvidas. Fazer apenas o interrogatório sobre uma fracção, às vezes mínima, da matéria, assenta numa exigência difícil de justificar, sobretudo no nosso sistema de ensino actual não individualizado¹. Considera-se que todos os alunos aprenderam e supõe-se que a qualidade das suas aprendizagens tenha sido homogénea. Perguntas incidindo arbitrariamente sobre uma parte do todo permitiriam assim elaborar um juízo válido.

Na realidade, se por pouca sorte o professor interroga sobre uma das poucas partes que o aluno não estudou, a vigésima parte do conjunto, por exemplo, é a catástrofe. Frequentemente também, uma simples distracção, uma fadiga passageira podem explicar uma má resposta num domínio bem conhecido.

Já vimos que, para prevenir esse perigo, os criadores de testes tentam em geral fazer ao menos três vezes a mesma pergunta, sob formas diferentes, o que não parece ser habitual nos exames tradicionais.

Além disso, a multiplicidade dos objectivos a atingir, como nos revelou o modelo de Guilford, dificilmente torna concebível um pequeno número de perguntas.

¹ Ver a parte 5.ª: O mito da curva de Gauss.

No entanto, de modo algum seria realista preconizar uma construção elaborada com todas as exigências para um pequeno exercício de controlo ou de diagnóstico.

Saber fazer as perguntas é possivelmente a capacidade mais necessária ao professor. Mas é uma arte difícil. Testemunham-nos os erros na escolha do nível de linguagem e as obscuridades que frequentemente se observam no enunciado dos exercícios de aplicação e das perguntas de exame.

Evidentemente que não existe receita universal para a redacção das perguntas. No entanto podemos utilizar como introdução algumas regras propostas por Thorndike¹.

1. Antes de começar a redigir uma pergunta, devemos ter presente qual o processo mental que desejamos que o aluno utilize para responder.
2. Iremos utilizar matérias novas ou uma apresentação nova das matérias nas perguntas feitas.
3. Começaremos as perguntas pelas palavras ou expressões seguintes: «Compare — Oponha — Dê as razões de... — Explique como... — Critique — Diga o que aconteceria se...» Não começar as perguntas por palavras como «O que? Quem? Quando? Cite...».
4. Devemos elaborar as perguntas de forma que sejam claras e precisas para cada aluno.
5. Uma pergunta sobre matéria controversa deve antes pedir argumentos em favor duma posição do que pedir uma simples tomada de posição.
6. Certifiquemo-nos de que a pergunta implica um comportamento que desejamos que na verdade seja adoptado pelo aluno.

¹ Segundo R. THORNDIKE e E. HAGEN, *Measurement and Evaluation in Psychology and Education*, Londres, Wiley, 1969, 3.^a ed.

7. Devemos adaptar a extensão e complexidade das perguntas ao nível de maturidade dos alunos.

A. Perguntas compreensíveis

Muitas respostas erradas não são devidas à ignorância da matéria, mas sim a malentendidos, a uma má compreensão das perguntas. O perigo aparece particularmente num domínio como o das matemáticas, porque nelas facilmente se combinam as três dificuldades: dificuldade inerente ao problema, dificuldade da linguagem abstracta das matemáticas e falhas linguísticas dos próprios matemáticos.

Mas nós bem sabemos que os historiadores, os geógrafos, os professores de ciências... ou de línguas não estão isentos de cair nas perguntas abstrusas.

Se os professores não se assegurarem da clareza das perguntas, correrão o risco de praticar uma injustiça social. A investigação contemporânea confirma que muitas crianças oriundas de meios socialmente menos favorecidos sofrem de graves desvantagens no domínio da linguagem. Davis e Haggard, por exemplo, demonstraram que basta modificar a forma dum problema, sem lhe mudar o sentido, para que a diferença de resultados entre crianças provenientes de meios socio-económicos favorecidos ou não vá de 12 a 32% ¹.

B. Ter em conta o nível de informação

Para verificar a capacidade de transferir as aprendizagens, é necessário muitas vezes situar os problemas num contexto não ainda evocado na aula. E é preciso ainda que o contexto tenha um sentido para o aluno.

¹ A. DAVIS, «Education for the Conservation of Human Resources», in *Progressive Education*, 1950, 27, pp. 221-224.

Certos problemazinhos, encontrados recentemente num manual, supunham o conhecimento das regras de jogo do ténis ou da maneira de preencher boletins de prognósticos do futebol. O que significarão aqui os fracassos, se os alunos não têm a possibilidade de se informarem do sentido real das perguntas, antes de tentarem responder-lhes?

C. Ensaiar ou experimentar previamente as perguntas

Tanto quanto possível, devemos ensaiar as perguntas antes de as utilizarmos num exame. Descobrimos assim as imprecisões, as falhas de redacção, os erros materiais e, também, o nível da dificuldade.

Um ensaio deste género é dificilmente realizável por um professor isolado. Ainda uma vez, o trabalho em grupo abre possibilidades neste domínio, pois que o exame foi preparado muito tempo antes.

D. Cálculo da facilidade das perguntas

Se for possível fazer testes prévios sobre um grande número de assuntos, esse cálculo far-se-á antes do exame. Se não, mesmo assim é útil proceder à operação, quer para compreender melhor a fisionomia das respostas, quer para uma utilização ulterior.

A *percentagem de bons resultados* é o índice mais simples. Recordemos, porém, que não é correcto calcular percentagens a partir de números baixos. Normalmente o número de assuntos deveria exceder a centena.

Uma vez conhecido o índice de facilidade, o professor saberá melhor como dosear o exame.

E. Cálculo da eficácia — Poder discriminativo

O índice de facilidade sozinho pode induzir em erro, porque não raro resulta de factores acidentais. É prudente, pois, acompanhá-lo do *índice de eficácia* que revela em que medida uma determinada pergunta *discrimina* os alunos bons dos fracos.

1. Método simples

Pidgeon e Yates propõem um método de cálculo simples, muito útil para quem ensina.

- * Dividir a aula em três grupos: superior, médio e inferior, na base dos *scores* totais no exame considerado.
- * Para cada *item*, ver qual a percentagem de cada um dos três grupos que obteve bom resultado.
- * Para cada *item*, a diferença entre a percentagem de bons resultados do grupo superior e a do grupo inferior permite uma boa avaliação da eficácia.

Se a diferença for fraca ou nula, significa que a pergunta não distingue bem os alunos bons dos fracos.

Veremos adiante que quanto mais apurada é a discriminação mais larga é a distribuição dos resultados (a curva de Gauss achata-se).

2. Método mais apurado

Agora um outro método, utilizado para o exame final da escola secundária em Inglaterra¹.

Formam-se quatro grupos A. B. C. D.

Dão-se as indicações seguintes:

- 1) Determinar, para cada um dos *items* e em cada um dos quatro grupos, o número de alunos que respondem correctamente.

¹ D. MATHER e al., «The C. S. E.», *A Handbook for Moderators*, Londres, Collins, 1965, p. 108.

- 2) Calcular a média obtida por cada um dos quatro grupos por cada *item* em questão.
- 3) Reunir todos esses dados num quadro de conjunto e calcular as médias para cada *item* nos quatro grupos. Se as quatro médias se ordenam da mesma maneira que os *scores* médios para o teste inteiro nos quatro grupos, pode-se afirmar que o *item* considerado contribui para a discriminação total do teste. Se não, o *item* é suspeito.

Exemplo

Pergunta n.º	1	2	3	4	5	6	7	Total
Maximo	6	6	6	6	10	8	8	50
Grupo A	4,9	5,1	5,3	5,2	4,3	5,8	4,3	34,9
Grupo B	3,2	4,9	5,8	4,1	4,7	5,9	3,8	32,3
Grupo C	3,4	4,6	5,4	3	3,4	3,6	3,9	27,3
Grupo D	2	3,4	5	2,8	2,3	4,1	2,3	21,8
Ordem dos scores médios	A	A	B	A	B	B	A	A
	C	B	C	B	A	A	C	B
	B	C	A	C	C	D	B	C
	D	D	D	D	D	C	D	D
Discrimina	Não	Sim	Não	Sim	Não	Não	Não	Sim

A resposta que deste modo se obtém é um tanto grosseira. Para os *items* considerados como aceitáveis, não podemos garantir se discriminam meticulosamente ou se, ao contrário, são apenas aceitáveis. Além disso, não sabemos bem como melhorar a redacção da pergunta ou a maneira de a avaliar.

Para obtermos informações mais precisas, as estatísticas utilizam técnicas complicadas¹. A solução gráfica seguinte é simples e contém os matizes desejáveis.

¹ Embora a utilização do ordenador as ponha hoje ao alcance de toda a gente.

Consideremos os resultados individuais de cada um dos 20 alunos de cada grupo. Tomemos por exemplo as perguntas 2, 3 e 4. Em vez de escrevermos os resultados em números, marcamos-los num quadro.

Pergunta 2—Máximo 6				
Scor. obt.	Grupos			
	A	B	C	D
6				
5				
4				
3				
2				
1				
0				

Pergunta 3—Máximo 6				
Scor. obt.	Grupos			
	A	B	C	D
6				
5				
4				
3				
2				
1				
0				

Pergunta 4—Máximo 6				
Scor. obt.	Grupos			
	A	B	C	D
6				
5				
4				
3				
2				
1				
0				

Se em quadros semelhantes, indicarmos o quartil superior, o médio e o inferior¹, aparecerá claramente o comportamento dos alunos.

Pergunta 2				
Pontos	A	B	C	D
6	QS	QS		
5	M	M	QS	QS
4	QI		M	
3		QI	QI	
2				M
1				QI
0				

Pergunta 3				
Pontos	A	B	C	D
6	QS	QS	QS	QS
5	M	M	M	M
4	QI	QI	QI	QI
3				
2				
1				
0				

Pergunta 4				
Pontos	A	B	C	D
6	QS	QS		QS
5	M	M	QS	
4	QI		M	
3		QI		M
2				
1			QI	
0				QI

¹ A MEDIANA: nota do meio numa série de notas ordenadas; se o número de notas é par, calcula-se a média aritmética entre as duas notas medianas

2 4 8 10 11 13

— 9 —

Numa escala de medição, a mediana é o ponto acima do qual se encontram metade dos casos exactamente.

○ QUARTIL SUPERIOR ou Q_3 é a nota do meio da metade superior da série.

○ QUARTIL INFERIOR ou Q_1 é a nota do meio da metade inferior.

Teoricamente, deveríamos observar uma descida contínua de A para D. Essa descida é mais ou menos acentuada de acordo com o poder discriminativo. Uma subida indica que um grupo inferior (segundo o resultado no conjunto do exame) obtém melhores notas que o grupo superior.

Observa-se:

Pergunta 2:

— Não há nenhuma subida; há uma tendência à descida mas não muito acentuada.

Item frouchamente discriminativo.

Pergunta 3:

— Nenhuma descida para o quartil superior.

— Subida para a mediana.

— Subida para o quartil inferior.

Item não discriminativo. De rejeitar.

Pergunta 4:

— Nenhuma subida.

— Descida bem marcada para a mediana.

— Descida muito acentuada para o quartil inferior.

Boa discriminação. Uma reserva, contudo, para o QS do grupo D: os professores que corrigiram foram muito generosos para os melhores alunos do grupo fraco (D).

II. Respostas abertas ou fechadas?

A uma pergunta de *resposta aberta*, o aluno responde espontaneamente, utilizando o seu próprio vocabulário. A *resposta diz-se fechada* se o aluno é levado a operar uma escolha entre várias respostas possíveis.

Tradicionalmente a escola limitou-se só ao primeiro tipo de perguntas. Desprezava, pois, um elemento de grande utilidade.

A. Respostas abertas (orais ou escritas)

São as perguntas mais naturais, aquelas que a cada instante fazemos na vida.

São particularmente convenientes quer para provas de controlo feitas rapidamente, no decurso do ensino e com toda a espontaneidade pelos professores, quer para verificação de aprendizagens por tal forma complexas que escapam à análise rigorosa.

Parecem pertencer ao seu âmbito a avaliação das capacidades superiores (criatividade, capacidade de julgar, espírito crítico, ...). Não exclusivamente, porém. Em primeiro lugar porque, como veremos em seguida, as perguntas de escolha múltipla bem construídas permitem explorações muito mais subtis do que parece. Em seguida, porque, com a preocupação da objectividade, do rigor, as pessoas que classificam tendem, de maneira inconsciente talvez, a reter das respostas abertas só os elementos mais concretos, mais factuais que elas contêm. Não será a divergência das avaliações função da subtilidade do objecto sobre o qual incidem?

«Deveremos admitir — escreve Vernon¹ — que os alunos muito bons têm ocasião de manifestar certas qualidades excepcionais no exame tradicional, e que um professor perspicaz pode às vezes aperceber-se delas, ao passo que os outros podem muito bem desclassificar a resposta por causa do seu anticonformismo». O debate está muito longe de estar encerrado.

Ao construir completamente a resposta, o aluno tenta provar duas coisas: o conhecimento da matéria e a capacidade verbal de expressão.

Durante muito tempo, julgou-se que as duas caminhavam a par. O velho adágio «Exprime-se claramente o que claramente se pensa» foi verdade aceite por muitos professores de nossos dias. E no entanto as coisas não são assim tão simples. A tradução verbal do pensamento não é mais que uma forma

¹ Bull. 4, p. 7.

de expressão entre outras, desprezadas durante séculos pelas classes sociais privilegiadas a quem repugnava qualquer trabalho manual, qualquer tratamento do real.

Renunciar por completo às perguntas com respostas abertas seria um erro numa civilização em que a comunicação verbal permanece dominante. Mas às vezes é bom isolar, ao menos parcialmente, os conhecimentos e as capacidades de expressão verbal — o que as respostas fechadas permitem. Os professores não devem hesitar em recorrer a elas, tanto mais que, como notam Pidgeon e Yates, nunca foi demonstrado que a sua utilização, mesmo frequente, prejudicasse o desenvolvimento verbal.

Uma resposta formulada em completa liberdade apresenta um outro inconveniente grave: o seu carácter único não se presta bem à avaliação por comparação com as respostas de outros indivíduos. É na verdade uma das principais causas dos desacordos entre as pessoas que corrigem, desacordos tantas vezes assinalados pela docimologia.

Notemos finalmente que, se as perguntas abertas se prestam particularmente à revelação de processos mentais superiores, os examinadores estão longe de pôr em prática essa possibilidade.

Krumm e Seidel¹ analisaram 2825 exercícios feitos na aula e 3827 perguntas abertas de exames. Diziam respeito todas ao ensino comercial em todas as Länder da Alemanha federal. O apelo à memória («conhecimentos» segundo Bloom) era respectivamente de 96,6% e de 93%...

B. Respostas fechadas. Perguntas (*items*) de escolha múltipla

1. Utilidade

O exame intensivo pretende verificar em pormenor a qualidade das aquisições. Apresenta não raro uma função diagnóstica. O exame extensivo aplica-se a uma matéria vasta.

¹ V. KRUMM e HEDEL, *Wirtschaftslehrtest*, Weinhein, 1970, citado por INGENKAMP Ed., *Tests in der Schulpraxis*, Weinhein, Beltz, 1972, p. 112.

Em ambos os casos, são necessárias numerosas verificações. Daí o recurso a perguntas de «respostas fechadas», sobretudo do tipo «de escolha múltipla».

Exemplo: o primeiro teste de inteligência utilizável na prática corrente foi construído por;

- a) Binet
- b) Galton
- c) Goddard
- d) Spearman
- e) Terman

2. Formar uma provisão de perguntas

Arranjar várias dezenas de perguntas precisas, para um único exame, submete a imaginação a uma rude prova. Constituir um ficheiro de *items* que se vai enriquecendo continuamente facilita a tarefa de forma considerável.

Põe-se apenas uma pergunta em cada ficha para poder assinalar também os «diversores», ou seja as respostas falsas, mas verosímeis, à medida que elas aparecem. Esses diversores são particularmente fornecidos pelos erros cometidos com mais frequência pelos alunos. Deste modo temos a certeza de respeitar a sua «lógica» e fica garantida a verosimilhança.

Os professores dum mesmo ramo podem facilmente unir os esforços nesta matéria.

Parece que nos encaminhamos para a criação de *banças de items*, centros especiais que podem pôr à disposição dos professores muitas centenas de perguntas de escolha múltipla, bem aferidas para determinadas zonas.

3. Explorar a gama das possibilidades lógicas

Responder a um *item* de escolha múltipla não é necessariamente um simples exercício de memória, muito longe disso ¹.

¹ Essa técnica foi mesmo aplicada com êxito à análise literária — ver B. CHOPPIN e A. PURVES, *A comparison of open-ended and multiple choice items dealing with literary understanding in Research in the Teaching of English*, 3, 1, 1969, pp. 15-24.

As investigações empreendidas pelos serviços do ensino superior, com vista à reforma dos exames de medicina em França, evidenciaram a riqueza do sistema ¹. É intencionalmente que escolhemos exemplos no domínio dos exames universitários, com a esperança de convencer os professores de que, seja qual for a disciplina que ensinam, não devem excluir *a priori* o exame objectivo.

No presente exame, distinguem-se oito tipos de perguntas.

a) Perguntas de complemento simples

O *item* apresenta-se, na intenção se bem que nem sempre na forma, como uma frase a completar.

Ex.: Entre os caracteres seguintes, o que se aplica a todas as enzimas é:

- A. todas contêm uma coenzima dissociável;
- B. são termoestáveis;
- C. contêm sempre azoto na sua molécula;
- D. contêm sempre fósforo na sua molécula;
- E. são dialisáveis.

Note-se que nem sempre é necessário que a resposta correcta figure entre os elementos propostos. C poderia substituir E e tornar-se: «Nenhuma das respostas precedentes».

O *item* pode tomar a forma negativa.

Ex.: Um hidrossol metálico tem todos os caracteres seguintes, excepto um. Indique qual.

A — B — C — D — E

Para responder, o candidato deve conhecer todos os caracteres do hidrossol. Todavia, a forma negativa obriga a uma ginástica intelectual que se sobrepõe à dificuldade inerente à matéria.

¹ Paris, Ministère de l'Éducation Nationale. Enseignement supérieur, Examens et concours, section médicale 1961. Todos os exemplos de tipo médico que vão seguir-se são extraídos dessa publicação.

b) Associação simples

Serve para verificar o conhecimento «de um certo número de entidades que podem ou não estar em relação».

Neste caso, teremos de fazer corresponder um elemento precedido duma letra a um elemento precedido dum algarismo (associação ou «emparelhamento»).

Ex.: *Ramificação posterior do nervo radial*

- A. Segmento de origem
 - B. Grupo das ramificações posteriores
 - C. Grupo das ramificações anteriores
 - D. Nervo interósseo posterior.
4. é aplicado na face posterior do ligamento interósseo
 6. contorna o colo do rádio
 7. inerva os músculos da camada superficial da região antebraquial posterior
 8. passa entre as duas extremidades do pequeno supinador
 9. enerva os músculos da camada profunda da região antibrachial posterior.

a) Associação composta

É uma simples variante da precedente

Exemplo

- A. Paludismo de plasmodium vivax
- B. Paludismo de plasmodium falciparum
- C. Ambos ao mesmo tempo (A e B)
- D. Nenhum dos dois.

Pergunta 1:

1. O desenvolvimento clínico tem todas as probabilidades de ser menos grave num homem de raça negra que num homem de raça branca (resposta B)

2. Uma associação de primaquina e de cloroquina é o tratamento específico para um ataque agudo (resposta A)
3. Os episódios clínicos suprimem-se por ingestão de cloroquina uma vez por semana em zona endémica (resposta C)
4. Cura dum a maneira definitiva pelo tratamento com cloroquina (resposta B)
5. Evita-se a contaminação, ingerindo cloroquina uma vez por semana (resposta D)

Note-se que o número de perguntas pode ser aumentado ou diminuído, segundo a importância dada à matéria.

d) Associação com exclusão de termo

No *item* seguinte, 4 dos 5 fenómenos enumerados são comuns a uma das perturbações A.B.C. Deve-se indicar a perturbação (A) e o fenómeno que não lhe corresponde (2).

Exemplo:

- A. *eosinofilia de importância diagnóstica*
- B. plasmocitose de importância diagnóstica
- C. linfocitose de importância diagnóstica
- 1.º triquinose
- 2.º mieloma múltiplo
- 3.º síndrome de Loeffler
- 4.º doença de Hodgkin
- 5.º xistosomiase

e) Análise de relação de causa a efeito

Ex.: A articulação cubito-radial superior permite movimentos de rotação limitados PORQUE a cabeça do rádio é envolvida pelo ligamento anular.

- A. A constatação e a razão proposta são ambas verdadeiras e existe entre as duas uma relação de causa e efeito.
- B. A constatação e a razão proposta são ambas verdadeiras e não há entre elas relação de causa e efeito.

- C. A constatação é verdadeira, mas a razão proposta é falsa.
- D. A constatação é falsa, mas a razão proposta é um facto ou um princípio aceite.
- E. A constatação e a razão proposta são ambas falsas.

f) **Análise de observação**

O conjunto complexo seguinte põe o candidato numa situação comparável à experiência real.

Exposição da doença. O doente é um homem de 21 anos que se queixa de mal-estar, de tosse, de febre.

A doença começou dez dias antes da admissão, por um malestar e uma tosse sem expectoração, manifestando-se nas 24 horas seguintes uma temperatura que variava entre os 37,8 e 38,3 que persistiu até ao momento da admissão.

Ao quarto dia de doença, a tosse acentuou-se, produzindo pequenas quantidades de expectoração branca e viscosa.

Três dias antes da admissão, começaram certos acessos paroxísticos de tosse, seguidos por vezes de vômitos. Notaram-se sensações de arrepios mas não verdadeiros arrepios com tremuras. Existe desde o quinto dia de doença uma dor paraesternal anterior à tosse.

Ao exame físico, temperatura 38,3, pulso 110, ritmo respiratório 32, tensão máxima $10\frac{1}{2}$, mínima 8.

O doente mostra-se bem desenvolvido, sem magreza, a doença parece aguda, sofre de dispneia mas não está cianosado.

O exame físico da caixa torácica revela vibrações vocais à palpação e à auscultação. O murmúrio vesicular é normal. Na axila esquerda ouvem-se ralos finos e a qualidade brônquica do som está aumentada, embora de intensidade normal.

A fórmula sanguínea é a seguinte: glóbulos brancos 3400 (polinucleares 30%, linfócitos 62%, monócitos 5%, eosinófilos 3%).

A radiografia do tórax revela um aumento da densidade da região peri-hilar com áreas mal definidas de densidade desigual, nebulosa nas duas bases e num campo pulmonar superior esquerdo.

Perguntas

1. *Qual é o diagnóstico mais provável?*
 - a) tuberculose
 - b) pneumonia com pneumococcus
 - c) pneumonia (primária típica) com virus
 - d) coccidiomicose
 - e) bronco-pneumonia

2. *Quais as manifestações físicas que provavelmente aparecem?*
 - a) esplomegália
 - b) manifestação de sofrimento meníngeo
 - c) atrito pleural
 - d) mudanças frequentes na distribuição dos sintomas torácicos
 - e) condensação lobar esquerda

3. *Qual dos exames de laboratório seguintes acompanha o diagnóstico?*
 - a) elevação e aumento das aglutinimas frias
 - b) hemocultura positiva
 - c) leucocitose marcada no início da convalescença
 - d) exame da expectoração
 - e) cutis-reacção positiva

4. *Que terapêutica deverá aplicar-se?*
 - a) repouso na cama e estreptomomicina
 - b) repouso na cama e penicilina
 - c) estreptomomicina e ácido para-amino-salicílico
 - d) repouso na cama e aureomicina
 - e) psicoterápia e reeducação física

5. Qual a evolução provável desta doença sem tratamento?

- a) a febre irá desaparecer espontaneamente com uma crise terminal
- b) a convalescença será progressiva com uma recaída previsível
- c) irá desenvolver-se um empiema residual
- d) irá aparecer, depois da cura, uma fibrosa residual
- e) poderá aparecer uma caverna pulmonar

g) **Comparações quantitativas**

Compare X a Y

X — pressão mecânica no capilar venoso

Y — pressão oncótica no capilar venoso

e diga se:

- A. X é maior que Y
- B. Y é maior que X
- C. X é igual a Y

h) **Relações**

Ou seja:

- 1. débito circulatório cutâneo — ET;
- 2. a quantidade de calor perdida por unidade de tempo;
- A. o aumento do primeiro é acompanhado do aumento do segundo; ou a diminuição do primeiro é acompanhada duma diminuição do segundo;
- B. o aumento do primeiro é acompanhado do aumento do segundo, ou a diminuição do primeiro é acompanhada do aumento do segundo;
- C. as variações do segundo são independentes das variações do primeiro.

i) Complementos agrupados

Processo utilizado quando uma pergunta pode ter mais que uma resposta correcta.

Ex.: cinco recrutas medem: 1,65 m - 1,67 m - 1,69 m - 1,63 m e 1,61 m.

1. A média das estaturas do grupo é 1,65 m.
 2. O desvio-padrão está próximo de 8
 3. O desvio-padrão está próximo de 2,8
 4. A amostragem tem boas probabilidades de nos informar sobre a estatura dos indivíduos em geral.
- A. 1, 2 e 3 são correctas
B. 1 e 3 são correctas
C. 2 e 4 são correctas
D. 4 é correcta
E. Só uma das proposições é correcta.

Pode-se ver, em apêndice, uma comparação entre um exame conduzido segundo o método tradicional e segundo o método de escolha múltipla.

4. Cálculo da eficácia dos diversores

Numa pergunta de escolha múltipla, importa verificar se os diversores desempenham bem o seu papel. Para isso, calcula-se, para cada pergunta, qual a percentagem de escolhas que cada possibilidade de resposta obteve.

a) Situação ideal

A	B	C	D	E
10%	10%	10%	60%	10%

D é a resposta boa.

As outras escolhas foram igualmente atractivas.

b) Situação a corrigir

A	B	C	D	E
3%	0%	35%	60%	2%

Apenas o diversor C operou realmente. Os outros não entrevistaram.

5. *Críticas e refutação parcial*

Embora o seu valor esteja há muito tempo provado, o exame mediante perguntas de escolha múltipla continua a suscitar vivas críticas. Algumas poderão facilmente ser refutadas; outras parecem justificadas; outras ainda nem podem ser rejeitadas nem confirmadas, por falta de critérios científicos.

a) **Uma objectividade enganadora**

A avaliação das respostas às perguntas de escolha múltipla é indiscutivelmente objectiva. As respostas correctas são definidas de antemão. Os alunos ou as encontram ou não. São evidentes as vantagens deste método.

Todavia a subjectividade está longe de ser eliminada. Subsiste, pelo menos em parte, na redacção das perguntas e na decisão respeitante à resposta a considerar como correcta.

Em primeiro lugar, na redacção das perguntas. Estas são o fruto da reflexão, da invenção dos examinadores que, em última análise, operam subjectivamente. Ao lado das perguntas em que pensaram, quantas outras, talvez mais válidas, não poderiam ter sido imaginadas?

Depois, na redacção das respostas consideradas ideais. Raramente se trata de respostas tão evidentemente correctas como 2 e 2 são 4. No decorrer da construção de testes com os nossos estudantes, encontrámos várias vezes respostas propostas como correctas, escapadas vitoriosamente através do crivo da análise matemática e que, no entanto, eram quanto muito grosseiras aproximações expressas em linguagem duvidosa.

Como nota Vernon, não sem malícia, a superioridade indiscutível dos exames «objectivos» sobre as provas tradicionais é devida com mais frequência a uma cuidadosa preparação

e à união de esforços e de competência de numerosos professores e psicopedagogos do que à própria natureza do exame¹.

Também se esquece que, entre a utopia da objectividade e a subjectividade total, existem muitos estádios intermediários.

b) Escolhas «correctas» contestáveis

Também este ponto foi excelentemente discutido por Vernon², que escreveu:

«É frequente que pessoas superiormente educadas acolhem perguntas de escolha múltipla como críticas tais como «Este item é estúpido» ou pretendem que essas escolhas, consideradas como «falsas» pelo organizador do teste, são tão admissíveis, senão mais, como a resposta dita «boa». Esta crítica deriva principalmente do facto de que o item objectivo não se baseia nas mesmas capacidades das perguntas tradicionais. Não há dúvida que exames construídos por amadores podem conter numerosos *items* maus. Também se encontram nas provas organizadas por profissionais, mas menos frequentemente, porque os *items* insuficientes ou são rejeitados no momento do controlo preliminar por especialistas da respectiva disciplina ou serão eliminados na altura da análise matemática das respostas (...). Além disso, os críticos correm o risco de esquecer que as suas reacções sofisticadas podem diferir profundamente das dos alunos em que foram ensaiados os *items*. Enfim, os críticos podem ler nas perguntas coisas que nem sequer vêm à mente de alunos inteligentes.»

c) Uma lotaria...

Numa pergunta fechada com a escolha entre duas respostas, uma correcta e outra falsa, a probabilidade de êxito do aluno

¹ VERNON, The C. S. E.: *An Introduction to Objective-Type Examinations*, Londres, H. M. S. O., 1964, pp. 4-5.

² VERNON, *o. c.*, p. 6.

que age completamente ao acaso é, teoricamente, de 50%. Os criadores de testes há muito tempo que o sabem.

Actualmente, recorre-se a maior parte das vezes a perguntas com cinco possibilidades de escolha, das quais uma é correcta e as outras apenas verosímeis (diversor)¹. Deste modo o factor acaso é consideravelmente reduzido. O comportamento dos alunos variará, aliás, de acordo com o seu conhecimento da matéria. Em virtude do cálculo das probabilidades, o estudante que ignora tudo terá uma reduzidíssima possibilidade de cair às cegas na solução correcta. Ao contrário um conhecimento parcial permite eliminar pela reflexão um certo número de diversores. A probabilidade de êxito, escolhendo às cegas dentre as possibilidades restantes, é então muito maior do que no caso precedente, e é justo.

Aliás, é possível reduzir largamente o papel do acaso, corrigindo os resultados obtidos mediante uma fórmula simples² que, no entanto, penaliza de forma exagerada, visto a parte da escolha às cegas variar segundo os alunos.

d) **Acrobacia mental**

Acusam-se os exames tradicionais de darem grande importância à facilidade de expressão. Podemos no entanto perguntar se os *items* de escolha múltipla um pouco complicados (já referimos exemplos) não apresentam um inconveniente pelo menos tão grave, juntando à dificuldade inerente à matéria em questão, a obrigação de destrinçar negações duplas, de apreender subtilidades lógicas ou, de forma mais geral, beneficiando uma certa aptidão para a abstracção a partir de dados

¹ As vezes ainda, convêm várias escolhas, mas em graus diferentes. Nesse caso, deve-se seleccionar a mais adequada.

² Número de respostas correctas — $\frac{\text{Número de respostas incorrectas}}{\text{Número de escolhas} - 1}$

verbais. P. Vernon verifica, em todo o caso, que a compreensão da leitura desempenha um papel importante no êxito dessas provas ¹.

Importa conhecer não apenas o nível de desenvolvimento mental dos alunos, mas ainda saber claramente o que se pretende: verificar o conhecimento da matéria, a capacidade de compreensão da leitura ou do raciocínio ou a combinação dos dois.

Não se compreende como seria possível a sondagem das qualidades intelectuais superiores sem um acréscimo da complexidade dos *items*. É a razão porque muitos autores julgam que as perguntas abertas continuarão a desempenhar um papel importante nos exames de nível mais avançado.

Em caso de recurso à correcção automática, surge uma outra dificuldade: a utilização de cartões de respostas em que geralmente apenas figuram os números das perguntas e as letras A, B, C, D, E, representando as cinco opções. Este sistema exige um suplemento de atenção por parte do aluno, particularmente quando não pode responder a certas perguntas e passa à frente. Se esquecer de passar também a linha correspondente no cartão das respostas, as consequências podem ser desastrosas. Todavia não se deve exagerar a dificuldade: Remmers, Gage e Rummel mostraram experimentalmente que o sistema é utilizável com crianças, a partir dos nove ou dez anos ².

Seja como for, os alunos deverão ser cuidadosamente treinados na técnica dos exames de escolha múltipla antes de se sujeitarem a uma prova decisiva. Com esta precaução e uma construção rigorosa, as perguntas de escolha múltipla dão bons resultados. Prova-o suficientemente a utilização intensiva que delas se faz nos países anglo-saxões, há já muitos anos.

¹ P. VERNON, «The Determinants of Reading Comprehension», in *Educational Psychological Measurement*, 1962, 22, pp. 269-286.

² H. REMMERS, N. GAGE e RUMMEL, *Educational Measurement and Evaluation*, New York, Harper, 1955, p. 246.

e) **Inconvenientes incertos**

Classificamos de incertas as seguintes acusações, porque, tanto quanto nos é dado saber, nenhuma investigação científica rigorosa lhes estabeleceu uma fundamentação segura. Parece que certas críticas só se justificam na medida em que o exame é mal organizado.

- * É mais fácil escolher, de entre outras, a resposta boa do que construí-la. A memória intervém demasiado. Corre-se o risco de que uma certa preguiça intelectual se instale, uma repugnância pelo esforço necessário, pela formulação clara do pensamento.
- * Sabendo que, para fazer o exame de língua materna, bastaria sublinhar ou indicar a resposta escolhida com uma cruz, os professores negligenciariam os exercícios de expressão.
- * Dar-se-ia a preferência a numerosos exercíciuzinhos a realizar em tempos muito breves. E assim não se prestaria atenção aos problemas que exigem longa reflexão. Ora, os estudos superiores e a vida real não nos poupam a esse género de provas.

C. **A modo de conclusão: um compromisso**

Parece que um compromisso se impõe, a dois níveis pelo menos.

Visto que perguntas abertas e perguntas fechadas parecem possuir cada uma vantagens particulares e pôr em causa aprendizagens diferentes, não vemos qualquer razão para reivindicar um monopólio para nenhuma das duas. Podem perfeitamente coexistir. Os melhores docimologistas são contudo de opinião que não será de misturar os dois tipos de perguntas numa mesma prova.

Aliás, existe uma fórmula intermediária entre perguntas abertas e fechadas e que cada vez mais irá impor-se, de acordo com os progressos da investigação pedagógica.

O exame consistiria num número bastante elevado de perguntas *abertas*, de alcance relativamente limitado, previamente aferidas. Seria pois conhecido de antemão o leque das respostas prováveis, o que permitiria propor um esquema de classificação muito aproximado do rigor da pergunta de resposta fechada.

Esta possibilidade já foi confirmada por numerosos trabalhos experimentais. Deste modo foi possível obter correlações quase perfeitas (.98 e .99) entre sete grupos de classificadores, representando cada um deles uma comissão de exames da Grã-Bretanha. A experiência foi feita com provas de história de nível superior do G. G. E. (fim do ensino secundário)¹.

Sem irem tão longe, os professores podem perfeitamente entender-se sobre os pontos a exigir nas respostas a perguntas abertas.

Por outro lado, a maneira de redigir as perguntas pode fechar parcialmente as respostas e diminuir, portanto, o papel da subjectividade. Uma pergunta como: Escreva um começo apropriado à frase «...quando começou a chover» está, de algum modo, a meio caminho entre a prova subjectiva e a prova objectiva. Neste caso, os desacordos entre os classificadores convidados a pronunciar-se sobre a correcção das respostas serão bastante raros.

III. Subjectividade-Objectividade

A. Teoria

O facto de nos apoiarmos numa taxonomia em vez de nos abandonarmos à simples inspiração para redigir as perguntas reduz o jogo da subjectividade. Todavia está longe de ser totalmente eliminada.

¹ JOINT MATRICULATION BOARD, *The Marking of Scripts in Advanced Level History*, Universities of Manchester, Liverpool, Leeds, Sheffield and Birmingham, 1964

Vimos que o criador de testes de conhecimentos, para saber que tipo de perguntas deve fazer, se baseia normalmente num quadro de dupla entrada: objectivos-matérias, em que cada intersecção sugere um tipo de *items*.

A subjectividade vai actuar primeiramente na interpretação das categorias taxonómicas, raras vezes definidas em termos de categorias observáveis. O redactor das perguntas, por exemplo, julga que tal *item* diz respeito à memória pura e simples, ao passo que para outro já tem que ver com a aplicação, com a análise, com a síntese, etc. Mas como saber se é assim na verdade?

O problema é muito mais complicado se pensarmos que uma tarefa complexa pode implicar a utilização de diversos processos que não pertencem ao mesmo nível taxonómico. Além disso, o nível taxonómico varia de acordo com as aprendizagens anteriores realizadas por diferentes indivíduos. Uma mesma pergunta exige uma síntese para um, isto é, no sentido bloomiano, uma produção original, divergente, ao passo que para outro exige apenas o recordar duma solução encontrada anteriormente e já utilizada várias vezes.

De mais a mais a pergunta surge aqui também da inspiração do seu redactor. Para a mesma matéria e ao mesmo nível taxonómico, é claro que poderíamos formular muitas outras perguntas, com a mesma dificuldade, mais fáceis, mais difíceis. Seja como for, nunca o redactor reúne todas as perguntas possíveis antes de fazer uma escolha (que poderia ser então uma amostragem representativa do conjunto).

Em suma, a introspecção, a inspiração continuam a desempenhar um grande papel, e o recurso a peritos, cujo comportamento não difere fundamentalmente, não resolve todo o problema, muito longe disso.

Para vencer esta dificuldade, J. R. Bormuth¹ propõe a operacionalização integral do método de construção das perguntas. A sua teoria ilustra bem os esforços de objecti-

¹ J. R. BORMUTH, *On the Theory of achievement test items*, Chicago, University Press, 1970.

vação empreendidos actualmente e, mesmo que não seja integralmente aplicável na prática, indica um caminho de investigação, e talvez um ideal para o qual devemos tender.

Para Bormuth, operacionalizar a redacção das perguntas é propor um conjunto de manipulações observáveis por todos, permitindo derivar do ensino um *item* de teste. «Uma definição *duma classe de itens de teste* de conhecimentos consiste numa série de directivas que indicam ao redactor como deverá redispôr segmentos do ensino para obter *itens* do tipo desejado. Essas directivas nunca devem fazer apelo à introspecção do redactor...»¹. E Bormuth continua: «No caso de o experimentador não poder verificar se os *itens* são precisamente do tipo que ele pretende e se outros experimentadores não podem construir *itens* que possam dar certeza dum mesmo tipo, esses outros experimentadores não podem de modo algum pretender refutar ou confirmar os resultados originais. Em casos desses, o estudo original também não tem valor»².

Bormuth propõe um modelo de definição operacional das perguntas.

1. Num primeiro tempo, dá-se uma estrutura sintáctica à matéria a explorar.
2. Em seguida, operam sobre essa sintaxe operações definidas e transformam os segmentos de ensino considerados em *itens* de testes.

W. Hively, H. Patterson e S. Page tomam uma posição³ próxima da de J. Bormuth. Apoiando-se sobre a teoria da «generalisabilidade» de L. Cronbach, definem um teste de conhecimentos como uma amostra tirada dum conjunto de *itens* bem definidos⁴. Para obter formas paralelas ao teste

¹ J. R. BORMUTH, *ibid.*, p. 5.

² *Idem.*

³ W. HIVELY, H. PATTERSON e S. PAGE, *A Universe defined system of arithmetic achievement tests*, in *Journal of Educational Measurement*, 5, 1968, 275-289.

⁴ Isto supõe todavia que o conjunto dos *itens* possíveis é finito, o que está longe de ser verdade para todos os ramos, pelo menos ao nível dos *itens* específicos.

inicial, basta, nesse caso, tirar ao acaso uma nova amostra de *items*.

O problema consiste evidentemente em definir o conjunto, o domínio. Deveria permiti-lo uma análise lógica ou psicológica (classes comportamentais).

W. Hively et al. distinguem:

- *A forma de items*, isto é regras, que permitem produzir um conjunto de *items* (= classe de *items* segundo Bormuth).
- *O universo de items*, isto é uma colecção de formas de *items*.

Ex.: O universo da subtracção dos números inteiros é o conjunto das formas de *items* que permitem explorar todos os aspectos.

— *A família de testes paralelos aleatórios*.

Conjunto de testes constituídos em função dum plano de amostragem aleatória, incidindo sobre um universo de *items*.

- Ex.: 1. Gere um *item* para cada forma.
2. Ordene os *items* ao acaso.
3. Recomece as duas operações precedentes tantas vezes quantos testes paralelos desejar.

O limite do sistema de construção operacional dos *items* proposto por Bormuth é claramente reconhecido por Diedrich¹. Com efeito, ele calculou que, para um manual de física com cerca de 16 000 frases, o conjunto de transformações daria 960 000 *items*. Nunca um organizador de testes percorrerá um semelhante caminho. É pois para uma posição intermediária que teremos de orientar-nos e, como escreve H. Rupprecht, «A formulação de *items* operacionalmente definidos não dispensa o construtor de testes de analisar os textos de ensino e de escolher as unidades a testar»¹. Não deixaremos de repetir que a inter-

¹ H. RUPPRECHT, *Konstruktion von Testaufgaben nach einem Verfahren von Bormuth*, in K. KLAUER et al., *Lernzielorientierte Tests*, Düsseldorf, Verlag Schwann, 1972.

venção reflectida, com o que ela inevitavelmente introduz de subjectividade, tenta compensar os limites práticos duma teoria.

Para evitar este retorno ao subjectivismo, poderemos tirar ao acaso as unidades a testar, em seguida aplicar os princípios de Bormuth.

Popham e a sua equipa adoptam uma posição muito mais próxima da prática escolar actual. Ajudado em cada um dos casos por um especialista pelo menos da matéria em causa, por vários docimólogos e professores que ensinam ao nível a que se destinam as perguntas, o grupo de Popham define em primeiro lugar os objectivos de conteúdo ou de atitudes relativamente gerais, depois divide-os em micro-objectivos. Em seguida aparece a redacção das perguntas cuja congruência com os objectivos é discutida em cada caso por um grupo de educadores, de professores e de investigadores.

Este processo é naturalmente menos rigoroso que o de Bormuth ou de Hively, mas oferece no entanto mais garantias do que a maneira subjectiva clássica e apresenta a vantagem de ser quase imediatamente realizável em todos os círculos escolares.

Pouco a pouco vão-se elaborando bancos de objectivos susceptíveis de prestar inestimáveis serviços à prática escolar, pelo menos no plano diagnóstico.

B. Alguns exemplos

1. O teste de estruturação perceptiva (*cloze test*)¹

Tradicionalmente, para testar a compreensão dum texto, o avaliador faz incidir um certo número de perguntas sobre

¹ N. do Trad. — A expressão «*test de closure*» usada em francês como equivalente do inglês *cloze test* refere-se a uma prova de estruturação perceptiva e que consiste em suprimir palavras num texto que o leitor terá de reconstituir mentalmente para apreender o sentido das frases. Trata-se, pois, da modificação de um conjunto perceptivo, da criação duma nova estrutura (ou forma, «gestalt»). Este teste permite, nomeadamente, avaliar o grau e a rapidez da compreensão em leitura. A palavra *closure* aparenta-se ao termo português *clausura* (ideia de fechar, de cercar).

o que julga necessário verificar (escolha subjectiva da matéria). Determina, subjectivamente também, o grau de dificuldade do conteúdo do problema posto e da forma utilizada para o exprimir. Por conseguinte, uma pergunta lógica ou formalmente fácil pode dizer respeito a um texto difícil, ao passo que uma pergunta difícil pode dizer respeito a um texto fácil.

O teste de «clausura» parece evitar esses escolhos. Consiste em suprimir uma palavra em cada cinco e em convidar o aluno a preencher as lacunas assim criadas. Numa primeira forma do teste, suprime-se, por exemplo, a 1.^a, a 6.^a e a 11.^a palavra, etc.; numa segunda forma, suprime-se a 2.^a, 7.^a e 12.^a palavras, etc.

Com cinco formas diferentes, ficam assim incluídas nas lacunas todas as palavras do texto. Obtiveram-se deste modo todas as perguntas possíveis nesta *forma de items*.

Esta prova, duma simplicidade de construção perfeita, permite uma boa avaliação da compreensão e da legibilidade.

2. *Teste de compreensão da leitura*

Quando mais não seja para completar as indicações fornecidas para o teste de «clausura», pode-se desejar construir um teste de compreensão, de contextura mais clássica.

Uma oposição entre a classificação abstracta de Davis e algumas regras operacionais inspiradas em Bormuth mostra bem, na nossa opinião, a diferença fundamental entre as duas maneiras de operar.

Exemplos (Davis): Matéria = compreensão da leitura.
Objectivos.

Nível I: Perguntas sobre o sentido duma palavra.

Nível II: Perguntas sobre a organização duma passagem.

Nível III: Perguntas sobre informações contidas no texto.

Nível IV: Perguntas sobre informações que devem ser inferidas a partir do texto.

Nível V: Perguntas sobre o ponto de vista do autor.

Neste exemplo, vemos que Davis propõe cinco níveis nos quais o redactor do texto vai situar subjectivamente as suas perguntas.

Bormuth inverte parcialmente a maneira de proceder no domínio dos objectivos. Tendo em conta os objectivos gerais, procuram-se regras de produção de *items*. Em seguida, uma vez encontradas aquelas, consideram-se os objectivos intermediários como, por exemplo, os da taxonomia de Bloom, para os quais podem servir.

Alguns exemplos directamente inspirados em P. Menzel¹.

Supunhamos que se quer testar a compreensão de *A Cigarra e a Formiga*. A escala de níveis de Davis é substituída por uma estruturação de base linguística que incidirá nomeadamente sobre:

a) **A Compreensão das palavras**

Poderia ser testada colhendo uma amostragem das palavras do texto que não pertencem a um vocabulário de base dado e construindo para cada uma o mesmo tipo de perguntas (fundada, por exemplo, sobre um conjunto de palavras que pertencem ao mesmo campo semântico).

- A nortada é: — a chuva?
- a neve?
- o vento?
- o nevoeiro?

b) **A compreensão da estrutura da frase**

Tira-se à sorte um certo número de frases e aplica-se uma mesma regra aos seus diversos constituintes.

- Quem se encontra muito desprovida?
- Quando é que a cigarra se encontra muito desprovida?
- Etc.

c) **A compreensão das relações anafóricas**

- Quem diz: «Eu pagar-vos-ei»?
- Quem é «esta emprestadora»?

d) **A compreensão das relações entre as frases e entre os parágrafos**

A sucessão temporal ou as relações causais, por exemplo, poderiam dar lugar a perguntas em que frases ou parágrafos apresentados em desordem devem ser reordenados.

¹ P. MENZEL, *The linguistic bases of the theory of writing items for instruction stated in natural language*, ANEXO a J. BORMUTH, *op. cit.*

De modo nenhum apresentamos estes exemplos como os melhores possíveis, mas procuramos simplesmente ilustrar uma direcção de investigação.

3. Formas de itens para a subtracção

No seu artigo já citado, W. Hively e outros propõem um programa completo de aritmética elementar, repartido por nove «universos» com a lista das formas de *itens*. Apresentamos para o universo «Subtracção dos números inteiros», alguns exemplos de formas acompanhadas das regras de geração.

Exemplos de formas de item para o universo da subtracção¹ *

Títulos descritivo	Exemplo de item	Forma geral	Regra de geração
Número grande > 10	$\begin{array}{r} 13 \\ - 6 \\ \hline \end{array}$	$\begin{array}{r} A \\ - B \\ \hline \end{array}$	<ol style="list-style-type: none"> $A = 1a$; $B = b$. $(a < b) \in U$. $\{ H, V \}$.
Um desconto, o pequeno número só tem um algarismo (1).	$\begin{array}{r} 53 \\ - 7 \\ \hline \end{array}$	$\begin{array}{r} A \\ - B \\ \hline \end{array}$	<ol style="list-style-type: none"> $A = a_1a_2$; $B = b$. $a_1 \in U - \{ 1 \}$ $(b > a_2) \in U_0$.
Um desconto do 0	$\begin{array}{r} 403 \\ - 138 \\ \hline \end{array}$	$\begin{array}{r} A \\ - B \\ \hline \end{array}$	<ol style="list-style-type: none"> $N \in \{ 3, 4 \}$. $A = a_1a_2\dots$; $B = b_1b_2\dots$ $(a_1 > b_1), (a_3 < b_3), (a_4 \geq b_4) \in U_0$. $b_2 \in U_0$. $a_2 = 0$. $P \{ \{ 1, 2, 3 \}, \{ 4 \} \}$
Equação: falta o termo subtrativo.	$42 - = 25$	$A - = B$	<ol style="list-style-type: none"> $A = a_1a_2$; $B = b_1b_2$. $a_1 \in U$. $a_2, b_1, b_2 \in U_0$. Verificar: $0 < B < A$.

¹ HIVELY W. et al., *o. c.*, pág. 281.

* *N. do Trad.* — O processo de subtracção ensinado nas escolas belgas é diferente do português — quando um algarismo do pequeno número não cabe no algarismo correspondente do grande número, *desconta-se* (ou *emprunte*) do algarismo da esquerda 10 unidades para tornar possível a subtracção. O processo português é o de *acrescento* ao algarismo à esquerda do algarismo que se subtrai.

Explicação da classificação

As letras maiúsculas **A, B, ...** representam números.

As letras minúsculas **a, b, ...** representam algarismos.

$(a < b) \in \{\dots\dots\dots\}$: escolher dois números ao acaso sem substituição
a deverá ser menor que **b**.

$\{H, V\}$: escolher uma disposição horizontal ou vertical.

$U = \{1, 2, \dots\dots\dots 9\}$.

$U_0 \square \{0, 1, \dots\dots\dots 9\}$.

N = número de algarismos em **A**.

N_A = número de algarismos em cada número do problema.

$P\{A, B, \dots\}$: escolher uma permuta dos elementos no conjunto (se o conjunto consiste em índices, permutar estes elementos indexes).

$x \in \{\dots\dots\dots\}$: escolher ao acaso um valor para **x** no conjunto dado.

a, b, c $\in \{\dots\dots\dots\} = a, b, c$ são escolhidos no conjunto dado com substituição (tiragem aleatória simples).

4. Exemplo do sistema de geração de itens para a matemática nova no início da escola primária¹.

Numa série de sete quadros iniciais, são sintetizadas as dimensões e os valores possíveis das principais unidades de matérias. Apresentamos uma dessas unidades.

III A. Estudo das qualidades de objectos e de conjuntos de objectos

Ponto de aplicação	Objectos isolados			B. Conjuntos de objectos		
	1	2	mais de 2	1	2	mais de 2
Número de propriedades encaradas simultaneamente						
Tipo de operação						
Descoberta das propriedades estatísticas (S) definidas						
positivamente						
negativamente						
Descoberta das propriedades dinâmicas (D)						

¹ J. PAQUAY-BECKERS, *D'une philosophie de la compensation à une pédagogie de la maîtrise*, Laboratoire de Pédagogie expérimentale de l'Université de Liège, 1973.

O quadro contém três dimensões:

- o ponto de aplicação da operação;
- o número de propriedades encaradas simultaneamente;
- o tipo de operação.

Paralelamente ao quadro centrado sobre a matéria, estabelece-se também o quadro dos comportamentos esperados dos alunos. Os comportamentos são determinados pelas características da situação-estímulo proposta. Fixam-se três dimensões:

- o canal que fornece a informação de base sobre a qual o aluno irá trabalhar (M);
- o canal em que o aluno é convidado a dar a sua resposta (E);
- o tipo de resposta esperada (O).

QUADRO IV: *Dimensões comportamentais*

Modo de apresentação: M / Modo de resposta: E				Objecto concreto	Desenho	Smbolo	Signo	Enunciado verbal	«Inventa»
		O		M ₁	M _{2a}	M _{2b}	M _{2c}	M ₃	
Objecto concreto E ₁	Escolha O ₁								
	I Produção: O ₂								
Desenho E _{2a}	Escolha: O ₁								
	Produção : O ₂								
Símbolo E _{2b}	Escolha : O ₁								
	Produção : O ₂								
Signo E _{2c}	Escolha: O ₁								
	Production : O ₂								
Enunciado verbal E ₃	Escolha : O ₁								
	Produção · O ₂								

Enfim, para as diferentes quadrículas do quadro III A, desenvolve-se a matriz de todos os exercícios possíveis, combinando dimensões comportamentais do quadro IV.

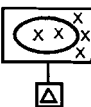
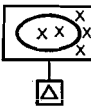
Consideremos, por exemplo, as quadrículas B-S do quadro III A (quadrículas tracejadas): estudo das propriedades estáticas de conjuntos de objectos, definidas positiva ou negativamente.


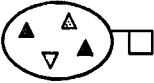

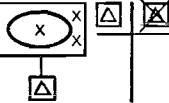

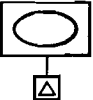
Cada célula da matriz assim obtida define uma classe de comportamentos equivalentes. Suprimem-se algumas células desta matriz ou por que descrevem um comportamento de simples cópia (ex. M_1E_1) ou porque se referem a um comportamento não introduzido por esse conteúdo (ex. utilização de sinais convencionais para o estudo das qualidades de conjuntos de objectos isolados).

Todas as quadrículas são exploradas pelos professores durante o seu ensino. Marcámos com um C as quadrículas que descrevem comportamentos susceptíveis de serem controlados por um teste colectivo papel-lápis. Os outros comportamentos devem ser controlados oralmente em relação a cada aluno individualmente. O professor constrói com esse fim quadros de resultados que vai preenchendo à medida que se desenrolam as actividades da turma.

OBJECTIVO IIa:

- Estudo das qualidades estáticas de conjuntos de objectos
- Uma só qualidade definida positiva ou negativamente.

Modo de apresentação / Modo de respostas E e O	Objectos concretos M ₁	Desenhos M _{2a}	Símbolos M _{2b}	Sinais convencionais M _{2c}	Enunciado verbal M ₃	Inventa: —
Objectos concretos: E ₁ Seleccção: O ₁ e produção: O ₂	Cópia	Cópia	 <p>— Escolha — entre esses blocos lógicos (O₁) ou — entre todo o seu material (O₂) os blocos (<i>concretos</i>) que podem ser ordenados nos lugares das cruzes</p>	Não introduzidos a este nível	<p>— Escolha entre este conjunto de blocos (<i>concretos</i>) o conjunto dos triângulos (O₁)</p> <p>— Construa um conjunto de triângulos (O₂)</p> <p>— Classifique esses blocos (<i>concretos</i>) segundo a cor O₁</p>	Inventa um conjunto que poderias construir com os blocos da nossa caixa (Dienes) (O ₁)
Desenhos: E _{2a} Seleccção: O ₁ e produção: O ₂	Cópia	Cópia	 <p>— Escolha — Entre esses desenhos de blocos (O₁) ou — Desenho (O₂) os blocos que podem ser ordenados no lugar das cruzes</p>	Não introduzidos a este nível	<p>— Escolha entre esses conjuntos de blocos</p> <p>— Desenhe um conjunto de triângulos</p> <p>— Classifique estes blocos (<i>desenhos</i>) segundo a cor O₁</p>	Inventa um conjunto e desenha-o (O ₂)

<p>Símbolos; E_{2b}</p> <p>Seleccção: O₁</p> <p>e produção: O₂</p>	<p>um conjunto de triângulos: objectos concretos (Dienes) (com um círculo à volta ou sob a forma de quadros)</p> <p>— Escolha entre estas etiquetas (O₁)</p>  <p>ou encontre (O₂) uma etiqueta representando cada conjunto</p> <p style="text-align: right;">C</p>	 <p>— Escolha de entre estas etiquetas (O₁)</p>  <p>procure (O₂) uma etiqueta representando cada conjunto.</p> <p style="text-align: right;">C</p>	 <p>— Ordene as peças do quadrado no diagrama ou vice-versa</p> <p style="text-align: right;">C</p>	<p>Não introduzidos a este nível</p>	<p>— Escolha de entre estas etiquetas (O₁) ou procure uma etiqueta (O₂) para representar o conjunto de todos os triângulos da nossa cozinha.</p> <p style="text-align: right;">C</p>	<p>— Procura uma etiqueta para representar o conjunto que tu investaste (O₂)</p> <p style="text-align: right;">C</p>
<p>Sinais convencionais E_{2a}</p> <p>Seleccção: O₂</p> <p>e produção: O₃</p>	<p>Não introduzidos a este nível</p>	<p>Não introduzidos a este nível</p>	<p>Não introduzidos a este nível</p>	<p>Não introduzidos a este nível</p>	<p>Não introduzidos a este nível</p>	<p>Não introduzidos a este nível</p>
<p>Enunciado verbal: E₃</p> <p>Seleccção: O₁</p> <p>e produção: O₂</p>	<p>um conjunto de triângulos: objectos concretos (Dienes) cercados por uma corda ou sob forma de quadros.</p> <p>— Descreva este conjunto (O₂)</p> <p>— Este pequeno quadrado vermelho pertence ao conjunto? Porquê? O₁</p>	 <p>Descreva este conjunto (O₂)</p> <p>Este pequeno quadrado vermelho pertence ao conjunto O₁? Porquê?</p>	 <p>— Descreva um objecto que pertença ou não ao conjunto (O₁)</p> <p>— Um quadrado vermelho pertence ao conjunto O₁? Porquê?</p>	<p>Não introduzidos a este nível</p>	<p>Cópia</p>	<p>— Inventa um conjunto e descreve-o (O₁)</p>

À primeira vista, parece difícil semelhante maneira de proceder. Todavia a experiência mostra que um professor bem iniciado no sistema acaba por considerá-lo como indispensável à prática esclarecida da sua actividade.

Não é menos verdade que, no contexto geral deste método de trabalho, a cooperação entre os centros de investigação e as escolas revela-se indispensável.

C. Conclusão

As operações muito «mecânicas» ou pelo menos muito analíticas que acabámos de considerar podem parecer muito afastadas do meio activo, duma pedagogia funcional. Para evitar qualquer malentendido, lembremos que o fim em vista é sobretudo sugerir instrumentos *diagnósticos*. Em nada nos informam sobre o valor da pedagogia utilizada. Seja ela qual for, exige um conteúdo e controlos de aquisição.

Poderíamos acantonar-nos nos dois primeiros exemplos simples, de entre aqueles que enumerámos. Mas não quisemos. Porque iríamos nós fazer supor que a pedagogia foge à complexidade crescente das outras disciplinas?

CAPÍTULO 3

A CLASSIFICAÇÃO

I. Um preâmbulo indispensável: a curva de Gauss

É impossível discutir a avaliação dos resultados sem buscar apoio nalgumas noções matemáticas. Não esqueceremos a promessa feita no início deste livro: bastará a aritmética elementar!

A. A curva de Gauss, imagem da probabilidade

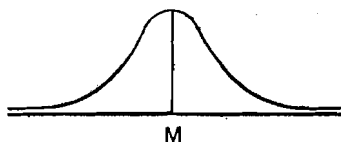
Um saco contém duas bolas, semelhantes em tudo excepto na cor: uma é vermelha, outra branca.

Nestas condições, cada bola tem uma probabilidade em duas de sair em cada tiragem ao acaso.

Se repetirmos este jogo poucas vezes, é possível que tiremos repetidamente a mesma cor. Mas à medida que aumentarmos o número de tiragens, essa probabilidade diminui.

Imaginemos 10 tiragens consecutivas. É muito provável que o vermelho venha a sair aproximadamente tantas vezes como o branco. Ao contrário, é muito pouco provável que se tire 100 vezes consecutivas a bola vermelha ou a bola branca. É já um bocadinho mais provável que se obtenha 99 vezes a vermelha e uma vez a branca; mais provável já, 98 vezes a vermelha e 2 vezes a branca, etc. Numa palavra, a probabilidade vai aumentar até 50 vermelhas e 50 brancas, em seguida diminuir progressivamente até chegar a uma vermelha e 99 brancas.

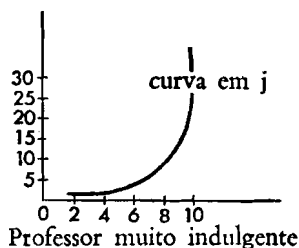
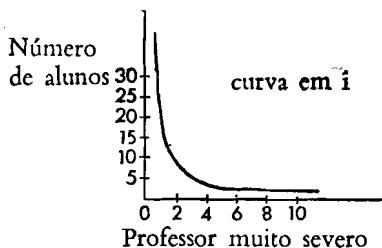
Num número elevado de tiragens, este movimento ascendente-descendente corresponde a uma curva com a forma de um sino. É a famosa curva de Gauss cujas metades são simétricas em relação à média aritmética e cujas extremidades nunca chegam a tocar a linha do zero, pois a probabilidade nula só existe no infinito.



Esta *distribuição*, chamada *normal* é a imagem de muitas qualidades humanas, tais como elas se repartem em grupos numerosos, *tornados ao acaso*. Deste modo, entre os habitantes duma grande cidade, os homens de estatura média são os mais numerosos, ao passo que os gigantes e os anões são muito raros. Entre estes dois extremos, a população distribui-se segundo a curva de Gauss.

Até o erro muitas vezes se submete à lei normal. Se efectuarmos um grande número de medições, veremos muito provavelmente aparecer um erro de grandeza média e uma distribuição que vai do erro ínfimo ao erro máximo. Veremos depois como esta observação é importante.

Desde que o acaso não intervenha, a distribuição modifica-se. Um professor pode, por exemplo, dar uma forte dominante de notas más e muito poucas boas ou ao contrário. É possível que se obtenha então uma das duas curvas características seguintes:



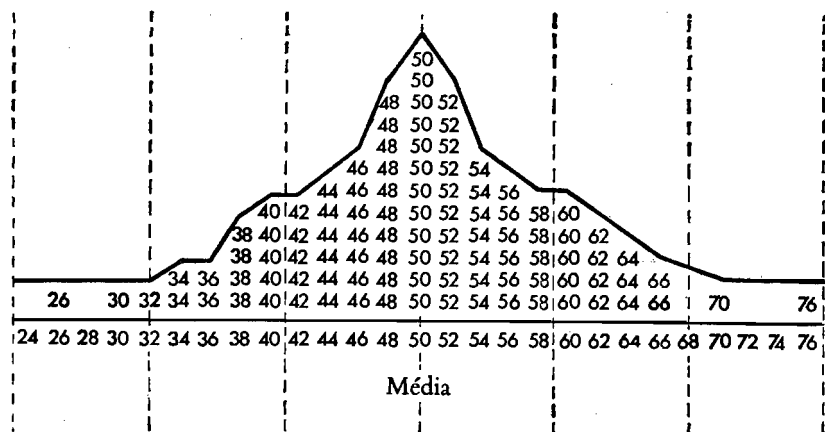
B. A curva de Gauss, imagem dos resultados do ensino não individualizado

Um professor que ensina de maneira não individualizada numa classe em que os alunos não são especialmente seleccionados, em geral dá ao seu curso um grau de dificuldade adaptado à maioria do grupo. Se o ajustamento é correcto, haverá muitos resultados médios, poucos muito bons e poucos muito maus. A distribuição desses resultados vai aproximar-se da curva gausseana.

Fenómeno semelhante se produz, mais espontaneamente ainda, nos exercícios que escapam à qualificação rigorosa, porque põem em jogo um conjunto complexo de factores.

Na figura seguinte dispusemos 100 resultados (imaginários) obtidos num exame, na escala de 100. Verifica-se que 12 alunos tiveram 50 (nota média), ao passo que apenas um teve a nota mais baixa (26) e apenas outro, a mais alta (76).

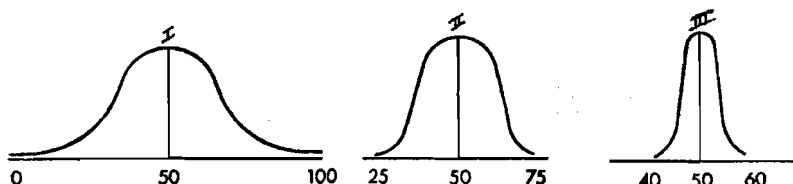
Observa-se igualmente que a linha correspondente à distribuição dos resultados não é uma curva perfeita em sino, mas apenas um esboço. Em casos semelhantes (já daremos algumas indicações a esse respeito), supõe-se que, se o número de sujeitos tivesse sido mais elevado, a curva ter-se-ia esbatido e considera-se a distribuição como «normal».



C. O desvio-padrão ou sigma, índice precioso

1. Significado

A variação dos 100 resultados que acabámos de examinar poderia ter sido ou maior, por exemplo de 0 a 100, ou mais pequena, por exemplo de 40 a 60, distribuindo-se sempre em curva de Gauss. Poderíamos ter obtido:



Nos três casos, a média é 50. Todavia trata-se de situações muito diferentes.

A margem de variação dos resultados é:

$$\text{Curva I: } 100 - 0 = 100$$

$$\text{Curva II: } 75 - 25 = 50$$

$$\text{Curva III: } 60 - 40 = 20$$

Sigma ou σ é um índice fácil de calcular (ver mais adiante o método de cálculo) que nos indica imediatamente como os resultados variam em relação à média aritmética das notas obtidas.

Para as 100 notas tomadas como exemplo, $\sigma = 9$. Se tomarmos um σ antes e depois da média, obteremos 41 e 59. Conte o número de resultados que se encontram entre estes dois limites e obterá 68 notas, ou seja neste caso 68% das notas de todo o grupo.

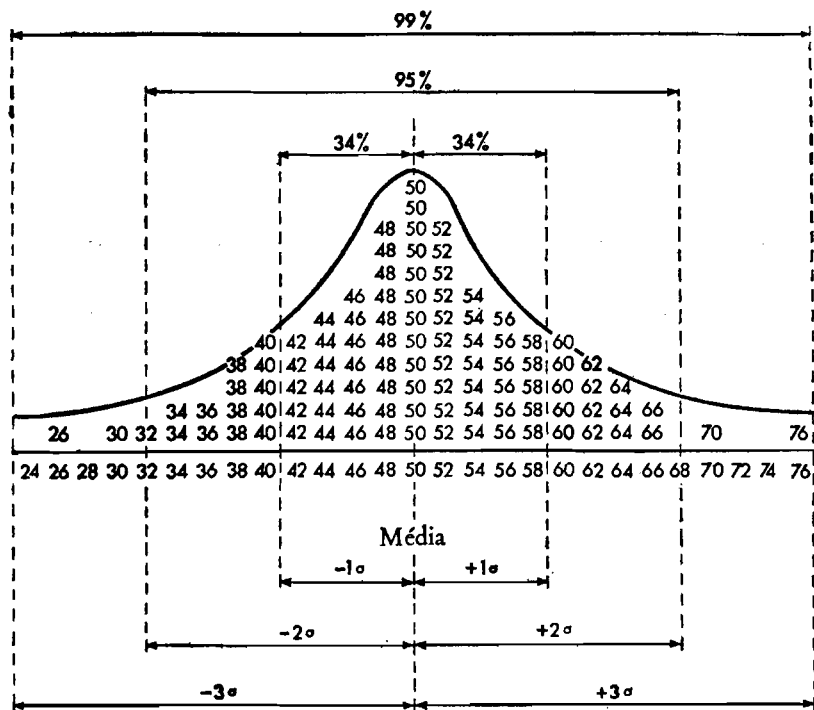
Acontece sempre assim numa *distribuição normal*. O σ indica sempre a mesma proporção dos resultados em relação ao conjunto.

1 σ de um lado e de outro da média = 68% das notas.

2 σ s de um lado e de outro da média = 95% das notas.

3 σ s de um lado e de outro da média = 99% das notas.

Vejam os agora a distribuição dos resultados com olhos diferentes:



Como a um sigma dado corresponde sempre a mesma superfície de área determinada pela curva, isto é a mesma proporção dos resultados, esse índice fornece-nos o meio de comparar os resultados atribuídos por professores diferentes, contanto que esses resultados se distribuam normalmente. Em breve, mostraremos como.

2. Avaliação rápida da média e do sigma

a) Problema

O cálculo rigoroso da média e do desvio-padrão (sigma) é relativamente difícil. E depois, quando a distribuição não é

normal, a média aritmética dá uma ideia falsa da tendência central.

Um processo simples permite evitar longos cálculos. Fornece apenas resultados *aproximativos*. Bastam, no entanto, em quase todos os casos da prática corrente na escola ¹.

Vejam as duas perguntas de exame às quais responderam dez alunos. São estas as notas atribuídas, as médias e os desvios-padrões calculados pelo método clássico.

<i>Aluno</i>	<i>Pergunta 1</i>	<i>Pergunta 2</i>
1	39	32
2	33	28
3	25	32
4	22	28
5	26	27
6	18	31
7	23	33
8	13	27
9	57	35
10	45	36
	301	309
	Média aritmética: 30	Média aritmética: 31
	Desvio-padrão: $\sigma = \sqrt{\frac{\sum d^2}{N}} = \sim 13.$	Desvio-padrão: $\sigma = \sim 3.$

¹ Segundo D. R. MATHER et al., *o. c.*

b) *Método simples de cálculo*

1.º Representação gráfica dos resultados

Notas	Pergunta 1	Pergunta 2
60		
58		
56	.	
54		
52		
50		
48		
46	.	
44		
42		
40		
38	.	
36		:
34	.	
32		..
30		
28		: :
26	:	
24		
22	:	
20		
18	.	
16		
14	.	
12		
10		

Margens de variação:

$$\text{Pergunta 1 : } 57 - 13 = 44$$

$$\text{Pergunta 2 : } 36 - 27 = 9$$

As notas da pergunta 1 são muito mais dispersas. Influenciam mais a classificação final que as da pergunta 2. Muitas vezes é útil ignorar as duas notas extremas. Neste caso:

$$\begin{aligned} \text{Margem na Pergunta 1} &= 27 \\ \text{na Pergunta 2} &= 8 \end{aligned}$$

Medianas

= notas do meio

Aqui, número par, portanto média entre 5.º e 6.º.

Pergunta 1:

$$(25 + 26) : 2 = 25,5$$

Pergunta 2:

$$(31 + 32) : 2 = 31,5$$

Para a pergunta 1, a diferença bastante nítida entre a média aritmética e a mediana (30 e 25,5) indica uma distribuição assimétrica das notas (aqui, maioria sob a mediana).

Para a pergunta 2, a simetria é boa (mediana 31 e média 31,5).

Note-se que, numa distribuição perfeitamente normal, a média e a mediana coincidem. Ver o exemplo.

2.º Avaliação do desvio-padrão (σ).

$\sigma = 3/4$ do desvio interquartil.

Portanto:

a. Procurar o quartil superior, isto é o meio das notas acima da mediana

Pergunta 1 = 39

Pergunta 2 = 33

b. Procurar o quartil inferior

Pergunta 1 = 22

Pergunta 2 = 28

c. Desvio interquartil

Pergunta 1 = 39 — 22 = 17

Pergunta 2 = 33 — 28 = 5

d. Avaliação do σ :

Pergunta 1 = $17 \times 3/4 = 12 3/4$

Pergunta 2 = $5 \times 3/4 = 3 3/4$

D. A concentração dos resultados à volta da média

Voltemos de novo à distribuição das 100 notas tomadas como exemplo (p. 137). Observemos que, quanto mais nos aproximamos da média, mais numerosas são as notas.

12 notas em 100 na média exacta

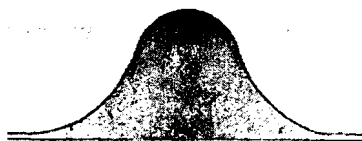
22 notas em 100 entre 48 e 52

68 notas em 100 entre 41 (-1σ) e 59 ($+1 \sigma$)

Imaginemos que se trate de composição francesa. Podendo o professor classificar de 0 a 100, um quarto mais ou menos das notas vai situar-se entre 48 e 52, ou seja a 4 pontos de distância.

O que significa uma diferença de 4 pontos em 100 no caso da dissertação francesa? É quase certo que, se fizermos avaliar o mesmo trabalho por dez professores diferentes, os desvios serão quase sempre superiores a 2. Ora, se é a média que constitui o limite do fracasso (aqui coincide com metade da escala total), podemos dizer, sem receio de nos enganarmos que, para 22% do grupo de alunos tomados como exemplo, tanto o fracasso como o êxito dependerão do puro acaso.

Vejamos em gradação uma imagem da densidade de população em relação à média:



E. Curva de Gauss preferida pelos professores

A curva de Gauss é um belíssimo instrumento de classificação, pois permite identificar os melhores (por exemplo, entre 25 e 35) e os menos bons.

No caso de um concurso em que numerosos candidatos pretendem um reduzido número de lugares, poderemos escolher, por exemplo, os que se situarem entre $+1\sigma$ e $+3\sigma$ (a parte que escapa à forte concentração à volta da média). Do mesmo modo, se desejarmos classificar os alunos — o que se fez de maneira bastante exclusiva no passado, mas que continua útil mesmo assim para permitir que cada um se situe —, será com certeza útil a distribuição *normal*.

Como já não se trata agora de grupos escolhidos ao acaso, mas de pessoas especialmente treinadas, o professor cria verticalmente as condições necessárias para uma distribuição gausseana. Como? Doseando os *items* segundo os índices de dificuldade (e de eficácia).

F. Como saber se uma distribuição é normal?

Tudo o que dissemos a propósito da curva de Gauss e do sigma só pode ser válido se os resultados se distribuírem

normalmente. Antes de iniciar qualquer operação, importa pois ter em conta se essa condição se verifica.

Existem processos matemáticos rigorosos para atingir esse objectivo, mas aqui continuaremos a contentar-nos com uma simples aproximação gráfica: o *histograma*.

O método é simples:

1. *Classificar os resultados*

Podemos ordená-los do mais elevado ao menos elevado ou inversamente.

Vejam os que resulta da classificação dos 100 resultados que tomámos como exemplos.

Notas	Número de alunos que obtiveram essas notas (frequências) (f)
26	1
30	1
32	1
34	2
36	2
38	4
40	5
42	5
44	6
46	7
48	10
50	12
52	10
54	7
56	6
58	5
60	5
62	4
64	3
66	2
70	1
76	1
	N = 100

Todavia, se as notas forem numerosas, este método não é nem rápido nem prático. É mais fácil formar classes.

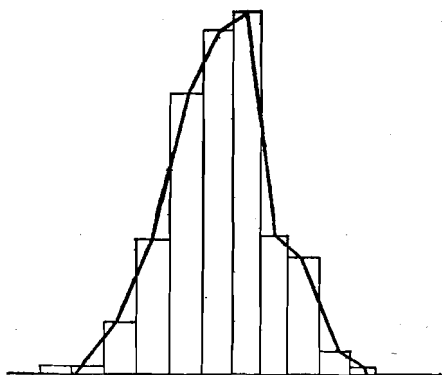
- 1.º Calcular a margem de variação entre as duas notas extremas: $76 - 26 = 50$.
- 2.º Dividir essa margem por 15¹: $50 : 15 = 3,33$.
- 3.º Escolher como intervalo de classe um dos dois números ímpares mais próximos: 3 ou 5. Neste caso, escolhe-se o 5, tendo em conta o reduzido número de notas.
- 4.º Pôr a nota superior no intervalo superior. Nota superior: 76; a classe superior, portanto, é: 74-75-76-77-78.

	<i>Classes</i>	<i>Frequências (f)</i>
1	74 - 78	I 1
2	69 - 73	I 1
3	64 - 68	III 5
4	59 - 63	III III 9
5	54 - 58	III III III 18
6	49 - 53	III III III III II 22
7	44 - 48	III III III III III 23
8	39 - 43	III III 10
9	34 - 38	III III 8
10	29 - 33	II 2
11	24 - 28	I 1
11 classes	Intervalo (i) = 5	Número de notas (N) = 100

2. *Desenhar o histograma*

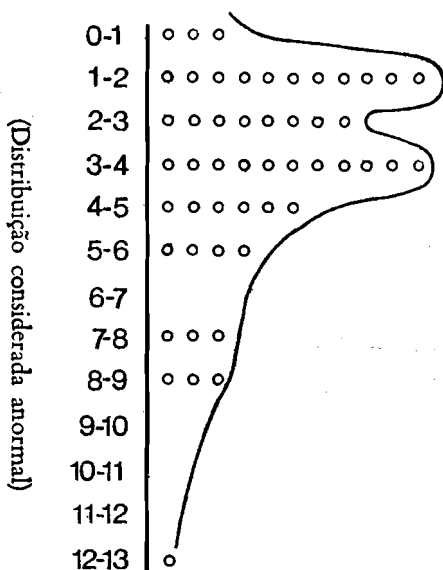
Na parte «frequências» do quadro acima, a marcação foi feita com muito cuidado. Basta manter a página horizontal para vermos aparecer o esqueleto do gráfico chamado histograma. Só resta traçar os rectângulos.

¹ Este número é arbitrário. A prática mostra que, na maior parte dos casos, ele conduz a uma boa distribuição.



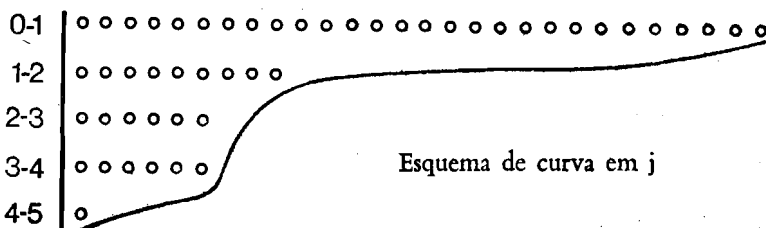
Se juntarmos os centros das bases superiores, obtém-se uma linha (polígono de frequências) que, neste caso, é um esboço com que nos contentaremos por considerarmos que a distribuição é provavelmente normal. Porquê? Porque o traçado só tem um *cimo*, mais ou menos *centrado*, a partir do qual as quantidades *vão decrescendo bastante regularmente, dum lado e doutro*.

Ao contrário, as duas distribuições seguintes *não* seriam consideradas *normais*.



Esquema de curva des centrada com dois cimos (curva bimodal)

(Distribuição considerada anormal)



II. Classificação subjectiva: escala de avaliação

A. Introdução

Por classificação subjectiva, entendemos toda a avaliação feita pelos professores duma maneira global seja imediatamente na altura da observação da *performance* dos alunos, seja depois de terem uma vista de conjunto dos *scores* particulares. De modo mais genérico, é subjectiva toda a nota a que se chega por qualquer processo não automatizável.

O grau de subjectividade varia consideravelmente. Ora actua quase sem limites (por exemplo: É bom — É muito bom...); ora quase se encaminha para um processo objectivo (exemplo: escala cujos escalões são definidos com uma precisão que praticamente elimina toda a interpretação pessoal).

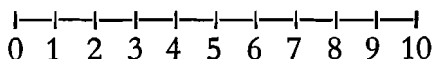
Em qualquer dos casos, a avaliação subjectiva faz-se por comparação a critérios externos ou internos. O professor que classifica conclui pela igualdade ou desigualdade em relação ao critério. Faz, portanto, uma classificação.

E assim podemos limitar esta parte ao estudo das escalas de avaliação.

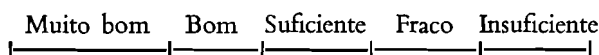
A escala de avaliação ajuda a *ordenar* indivíduos ou objectos em relação a uma qualidade que possuem num grau mais ou menos elevado.

Representa um *continuum* que teoricamente pode ir de um mínimo absoluto a um máximo absoluto, mas que, praticamente, quase nunca passa dum fracção, dum segmento desse absoluto. Do óptimo... ao... insuficiente, do pequenino... ao muito grande, etc.

A escala de avaliação, gráfica ou não, será
ou *numérica*:



ou *descritiva*:



A escala de avaliação é o instrumento mais frequentemente empregado na escola... e o mais mal conhecido. Podemos dizer, sem medo de nos enganarmos, que ela tem literalmente viciado uma grande parte dos exames a todos os níveis de escolaridade e terá provavelmente causado a pouca sorte dum considerável número de indivíduos.

Por essa razão, importa absolutamente conhecer a sua verdadeira natureza e sobretudo as suas imensas limitações.

B. Natureza e deficiências das escalas de avaliação

A escala de avaliação é *uma escala ordinal*

- 1) *Não tem nem zero nem máximo naturais ou absolutos.* Começa e acaba em qualquer número, à escolha do utilizador. Por consequência, mesmo que avaliem o mesmo fenómeno, do mesmo ponto de vista, nunca os juízes têm escalas rigorosamente idênticas. As diferenças podem ser consideráveis e são-no muitas vezes.

Exemplo: *Conhecimento do inglês, segunda língua*

Teoricamente, o zero poderia ser o momento em que o aluno não conhece ainda uma única palavra da língua. No entanto, raramente se avalia a partir desse ponto na escola.

Quanto ao conhecimento total, absoluto, evidentemente que não existe.

Imaginemos dois professores de inglês, avaliando, sem se combinarem, duas turmas em que todos os alunos fizeram uma curta exposição oral. Muito provavelmente, as duas extremidades das escalas de avaliação utilizadas independentemente pelos dois professores serão mais definidas pelo aluno considerado melhor e pelo aluno considerado mais fraco no seu grupo do que por uma avaliação abstracta da quantidade de inglês que deveria ser conhecida no nível pedagógico considerado.

- 2) *Os graus não são da mesma grandeza no interior duma mesma escala*

É impossível, por exemplo, demonstrar que a distância que separa um bom conhecimento do inglês dum conhecimento muito bom é a mesma que separa o conhecimento médio do bom.

Se os graus são A, B, C, D, E, não se pode dizer que $A - B = B - C = 0$

- 3) *Gráus correspondentes duma escala à outra (para uma mesma disciplina) não têm a mesma grandeza.*

Bom em inglês, para um professor, não é quantitativamente igual a um bom em inglês para um outro. Nada poderá estabelecer semelhante igualdade de forma matemática.

- 4) *Escalas referentes a disciplinas diferentes não se podem comparar*

Para os alunos duma mesma classe de escola primária em que um único professor ensina todas as disciplinas, ser bom em matemática não representa os mesmos fenómenos que ser bom em francês. Os escalões não poderão ser comparados no plano quantitativo.

De tudo isto resulta que as escalas de avaliação não permitem qualquer operação aritmética. Se classificarmos a composição

francesa e o inglês como segunda língua de 0 a 10, nada permite somar os dois resultados para avaliar um saber total. E no entanto é o que se faz, há dezenas de anos, nas escolas. Não é menos suspeito, senão o é ainda mais, utilizar escalas numéricas diferentes para as diversas disciplinas, em função duma ponderação empírica: de 0 a 120 para o Francês, de 0 a 80 para as Ciências, etc.

Se o sistema conseguiu funcionar, é porque, de qualquer modo, permitiu identificar os melhores (os que se classificam em tudo entre os primeiros) e os mais fracos (últimos em tudo).

C. Utilidade

Não obstante as limitações, a escala de avaliação é, em muitos casos, o único instrumento de que dispomos para concretizar o nosso juízo sobre comportamentos humanos complexos.

Efectivamente, permite trazer a um *dossier* escolar elementos *qualitativos*, tão objectivos quanto possível, *complementos indispensáveis* dos resultados *quantitativos* fornecidos pelos exames objectivos ou pelos testes.

D. Construção

Muitos educadores deixam-se iludir pela facilidade aparente com que se elabora uma escala.

Com efeito, a afinação desse instrumento não raro exige longos meses de trabalho, necessários à clarificação teórica e exige a utilização de técnicas refinadas como a análise factorial. Esta é notoriamente útil para determinar em que medida várias escalas avaliam as mesmas coisas sob designações diferentes.

Neste domínio também, importa por conseguinte que os professores possam apoiar-se o mais rapidamente possível em

serviços de investigação bem equipados com especialistas e ordenadores.

O que não quer dizer que os professores nada possam fazer, no entretanto.

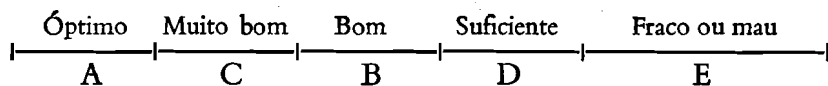
1. Quantos graus?

Teoricamente, uma escala de avaliação pode conter uma quantidade infinita de escalões. Na prática, os avaliadores não poderão classificar com qualquer segurança se não utilizarem um reduzido número de graus: 3, 5, 7 ou 9.

Tendo em conta a literatura experimental e múltiplas observações que pudemos fazer directamente, aconselhamos a que, *duma maneira geral, não se vá além de 5 escalões*. Uma escala dessas permite classificações relativamente seguras e fiéis, contanto que cada escalão seja muito claramente definido e os professores que classificam bem formados.

Praticamente:

- 1.º primeiro, classificam-se em três categorias: os melhores, os mais fracos, os restantes:
- 2.º depois, distribuem-se os «restantes» em bons, suficientes e fracos.



Note-se que o escalão C se apresenta no centro da escala, isto é na média aritmética. Como vimos a propósito da curva de Gauss, produz-se à volta desse ponto uma forte concentração. Escolher um número par de escalões criaria um centro na escala, à volta do qual oscilariam ao acaso muitos alunos.

Tratando-se duma distribuição normal, não se pode excluir que este veredicto do acaso afecte 20, 30 e até 40% dos alunos.

2. Definir o objecto da avaliação

Pedir, por exemplo, que se aprecie a «coragem» dos membros dum grupo dado — sem qualquer outra indicação — conduziria a respostas desprovidas de interesse. Que sentido ou sentidos terão dado os observadores à palavra «coragem»: ardor, vontade, zelo, perseverança, bravura, firmeza, estoicismo?

E mesmo se precisarmos que a palavra coragem significa firmeza diante do perigo, será fácil de estabelecer a distinção entre intrepidez e temeridade?

Para remediar imprecisões de classificação tais como *muito bom*, *suficiente*, *fraco*, etc., torna-se necessário juntar à escala uma descrição tão precisa quanto possível do traço ou do ponto a apreciar, e ilustrar a definição com situações tipo. Definir não com fórmulas abstractas mas com comportamentos precisos é uma condição *sine qua non* de validade.

Vejamos como Schonell¹ apresenta o traço «Confiança em si mesmo».

Confiança em si mesmo

Extremamente confiante em si.	Muito confiante nas suas próprias forças.	Confiante.	Falta de confiança. Tímido.	Falta extrema de confiança. Depende dos outros. Declina as responsabili- dades.
-------------------------------------	---	------------	-----------------------------------	---

Descrição do traço

«Na sua forma positiva, esta qualidade é assinalada pelas seguintes manifestações: o indivíduo conta consigo mesmo, é capaz de fazer face às dificuldades, sente-se seguro, é independente e está pronto a assumir responsabilidades.

¹ F. J. SCHONELL, *Backwardness in the Basic Subjects*, citado por F. WARBURNE, *Measurement of Personality* («Educational Research», Novembro 1961, p. 9).

«A criança que tem confiança em si mesma tenta avançar com o mínimo de assistência; a que a não tem deve ser constantemente ajudada. A primeira gosta de ver aquilo que é capaz de construir e produzir quando recebeu instruções claras; a segunda quer que lhe mastiguem a comida, que a ajudem durante toda a fase de realização».

Situações tipo

- 1.º Tem medo da escuridão?
- 2.º É capaz de cuidar de si mesmo e do que lhe pertence ou é preciso que esteja sempre alguém a seu lado?
- 3.º Viaja sozinho de eléctrico ou autocarro? (para crianças com mais de 9 anos).
- 4.º Fala livremente a visitas desconhecidas?
- 5.º É bom nos jogos? Sabe nadar?
- 6.º Está à vontade e responde com segurança aos exames orais?
- 7.º Lê bem, dramatiza bem um texto diante da turma?
- 8.º Lança-se com facilidade a tarefas novas ou faz continuamente perguntas aos companheiros e aos professores?

Definir os graus da escala

Exemplo: Organização do ensino da leitura

0	1	2	3	4	5
<i>Medíocre</i>	<i>Suficiente</i>	<i>Médio</i>	<i>Muito bom</i>	<i>Ótimo</i>	
Na leitura todos os alunos seguem a mesma progressão. Não trabalha em grupo.	Cf. 1. mas às vezes um aluno muito lento recebe menos de trabalho que os outros.	Formação de 2 ou 3 grupos, segundo as aptidões em leitura. Pouca flexibilidade no grupo.	Agrupamento segundo as aptidões. Flexibilidade.	Agrupamento depois de estudo aprofundado das aptidões das dificuldades encontradas. Grande flexibilidade.	

Como se pode notar, este último exemplo combina as escalas gráfica, numérica e descritiva.

Consequências duma definição insuficiente dos graus

Julgou-se que abandonar escalas numéricas (de cem pontos ou outras) por escalas literais (A. B. C. D. E) ou verbais (muito bem, bem, suficiente, fraco, insuficiente) se obteria uma melhor concordância entre as avaliações. Nada será possível, se os graus não forem bem definidos. Apresentamos, entre outros, três testemunhos experimentais desta afirmação..

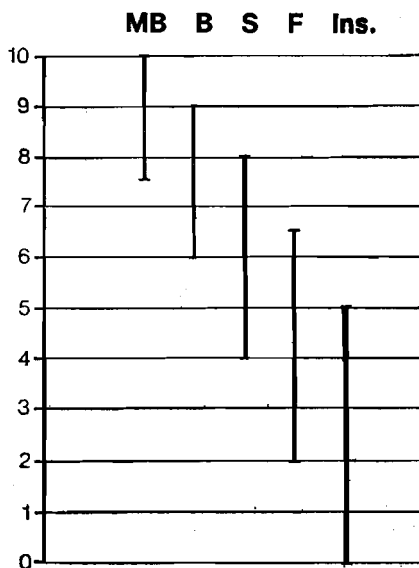
Por ocasião da introdução da escala verbal de cinco graus no ensino secundário renovado belga, os professores de francês e os professores de matemática¹ indicaram como procediam para classificar segundo o novo sistema. A maior parte atribuía mentalmente uma nota numérica na escala de dez e depois convertiam-na em avaliação verbal. Deste modo foi possível verificar em que medida as notas verbais correspondiam a avaliações concordantes.

Apresentamos primeiramente os extremos encontrados para uma mesma apreciação.

Muito bem:	limite inferior:	7,5 — 9
Bem:	limite superior:	7,5 — 9
	limite inferior:	6 — 7,5
Suficiente:	limite superior:	6 — 8
	limite inferior:	4 — 7
Fraco:	limite superior:	4 — 6,5
	limite inferior:	2 — 5
Insuficiente:	limite inferior:	2 — 5
	limite superior:	0 — 2

Torna-se mais elucidativa uma representação gráfica destas observações:

¹ *Inquérito sobre os modos de classificação*, Laboratório de Pedagogia experimental da Universidade de Liège, 1971 (documento policopiado).



Constata-se, portanto, que, segundo os professores:

- uma nota de 7,5 corresponde a muito bom, bom ou suficiente;
- uma nota de 6 corresponde a bom, suficiente ou fraco;
- uma nota de 4 corresponde a suficiente, fraco e insuficiente;
- uma nota de 2 corresponde a fraco ou insuficiente.

F. Bacher refere uma outra experiência também concludente¹. Demangeon e Larcebeau, a propósito dos mesmos trabalhos de francês e de cálculo, compararam a concordância entre professores conforme utilizavam as notas em algarismos ou quatro apreciações verbais: muito bom, bom, suficiente, fraco.

As discordâncias para um mesmo exercício são:

- para as notas em algarismos,
- de cerca de metade do máximo para o cálculo;
- de cerca de um quarto do máximo para o francês.

¹ M. DEMANGZON et S. LARCEBEAU, *Une expérience de correction multiple*, in BINOP, 1958, 14, 131-156. Citado por F. BACHER, *La docimologie*, o. c., p. 36.

— para as apreciações verbais:

- *Muito bom*, segundo os professores, é atribuído:
 - para o cálculo a 15% a 23% dos alunos;
 - para o francês a 0 a 30% dos alunos.
- O mesmo exemplar recebe a mesma apreciação:
 - para o cálculo por 42% dos que corrigem;
 - para o francês por 16% dos que corrigem.
- O mesmo exercício recebe: três apreciações diferentes
 - em 8% dos casos no cálculo;
 - em 20% dos casos no francês.

as quatro apreciações possíveis, em cálculo: em três dos casos.

Passar das notas em algarismos para as escalas A, B, C, D, E ou muito bem, bem, etc., não garante por si uma melhor avaliação.

Adoptando as escalas descritivas impostas pelo Ministério Austríaco da Educação Nacional, Weiss¹ convidou 92 professores a classificar duas composições de alunos do nível a que eles ensinavam (4.^a classe da primária) e 272 a classificar dois trabalhos de aritmética (problemas; 4.^a e 5.^a classes).

Os diagramas que seguem mostram como as notas são divergentes, para um mesmo trabalho (dispersão das notas em toda a extensão da escala).

CLASSIFICAÇÃO DE DUAS COMPOSIÇÕES

Classificação de duas redações

I		25%	50%	75%		Nota média	Sigma
Ortografia						2.89	1.04
Estilo						2.29	0.53
Fundo						2.08	0.79
Nota global						2.45	0.80

¹ R. WEISS, *op. cit.*

II

		25%	50%	75%		Nota média	Sigma
Ortografia	1	2	3	4	5	2.99	0.96
Estilo	1	2	3	4		2.12	1.05
Fundo	1	2	3	4		1.78	0.85
Nota global	1	2	3	4		2.54	0.80

Classificação de dois trabalhos de aritmética

		25%	50%	75%		Nota média	Sigma
I	1	2	3	4		2.56	0.77
II	2	3	4	5		3.55	0.80

Como se vê, não basta adoptar uma nova escala a substituir o sistema de notas tradicionais para que todos os problemas fiquem resolvidos. Muito longe disso.

E. Utilização

1) Quantos alunos por escalão?

Primeiro caso — Aluno comparado consigo mesmo

Os escalões marcam os graus a transpor na ascensão para um poder ou para um saber. O professor definiu um objectivo a atingir. Num ensino não individualizado, esse objectivo é comum à maioria, senão à totalidade dos alunos da turma. Deve ser escolhido evidentemente de forma que *todos* os alunos possam progredir nele. Será portanto um objectivo amplo, o que lhe pode prejudicar a precisão a ponto de o tornar inoperante.

É claro que procedemos actualmente como se, no início do ano escolar, todos os alunos se encontrassem aproximadamente no mesmo ponto da escala do saber. Supõe-se, portanto, que no ano anterior todos atingiram o grau superior da escala precedente.

Ora nós estamos convencidos que a realidade é diferente e a investigação confirma o nosso sentimento.

a) *Margens de variação da idade mental*

Aquando da revisão do teste de inteligência de Binet-Simon por Terman e Merrill (1937), constatou-se, nos Estados Unidos¹:

- que no 1.º ano da primária a idade mental variava de 4 a 8 anos;
- que no 6.º ano da primária, a idade mental ia de 8 a 16 anos;
- que no ensino secundário não era excepcional uma margem de variação de 8 a 10 anos.

b) *Margens de variação do rendimento escolar*

Basta observar como as normas dos testes de conhecimentos, escalonados por ano escolar, se sobrepõem, para tomarmos consciência da amplitude das margens de variação.

Apoiando-se nas suas próprias investigações e nas de Lindquist, Cornell, Learned e Wood, W.-W. Cook² fornece as indicações seguintes:

- Para a compreensão da leitura, para o vocabulário, ciências, geografia e história, a margem de variação do rendimento é:
 - * na primeira classe da primária, de 3 a 4 anos
 - * na quarta classe da primária, de 5 a 6 anos
 - * na sexta classe da primária, de 7 a 8 anos.
- Para o raciocínio aritmético e para o cálculo:
 - * na sexta classe da primária, de 6 a 7 anos.
- Em cultura geral (ciências, literatura, artes, história, geografia e língua materna), ao nível universitário:

¹ Cf. Q. McNEMAR, *The Revision of the Stanford-Binet Scale*, Boston, Houghton-M, 1942.

² W. W. COOK, *The Functions of measurement in the Facilitation of Learning*, in E. F. LINDQUIST, Ed., *Educational Measurement*, Washington, A. C. E., 1961, 4.ª ed., pp. 3-47.

- * 28% dos estudantes do 4.º ano são inferiores à média dos estudantes do 2.º ano;
- * 10% dos estudantes da 4.º ano são inferiores à média dos alunos do final do ensino secundário.

Semelhantes observações — que não se devem tomar à letra, mas sim ter em conta a sua tendência — poderiam multiplicar-se sem dificuldade.

Elas provam que, enquanto as nossas escolas funcionarem segundo o sistema das classes rígidas, a avaliação contínua acompanhada da preocupação de permitir a cada aluno uma progressão de acordo com o seu próprio andamento e com a possibilidade de ir sempre mais além, e mais além possível, será um logro.

Todavia, como por razões tecnológicas e económicas o ensino individualizado não poderá generalizar-se senão daqui por muitos anos só um ensino semi-individualizado oferece uma solução realista, imediatamente aplicável. Numa escola semi-individualizada, substituem-se à classe grupos homogêneos formados por ramos de disciplinas principais. Deste modo, um aluno dotado para uma língua estrangeira pode trabalhar a um nível A para essa disciplina, mas juntar-se ao grupo D para os problemas de aritmética ¹.

Para não tornar o texto pesado, damos em anexo uma descrição pormenorizada duma escola que funciona há cinquenta anos segundo este princípio.

Segundo caso: alunos comparados entre eles.

Mesmo que tentemos conduzir cada um dos alunos a um nível elevado, subsistirão provavelmente diferenças suficientes para fazer aparecer de novo uma distribuição gaussiana (ou pelo menos um esboço) em que a média apenas se deslocará para a parte superior.

¹ Não é por acaso que falamos de *problemas de aritmética* e não de aritmética em geral. Numa disciplina complexa como esta, um aluno pode ser bem dotado em cálculo, mas fraco em problemas. A investigação contemporânea mostra que, se os alunos não forem agrupados muito cuidadosamente, as margens de variação ficam consideráveis.

Parece, portanto, justificar-se a tentativa de povoar cada escalão de acordo com as percentagens que se aproximarem da distribuição normal.

Numa classe corrente, evidentemente que não é possível respeitar esta proporção de forma absoluta, em virtude do reduzido número de alunos. Teremos, por exemplo:

				A distribuição normal seria:
Ótimo	1	5% 2,5%
Muito bons	4	20% 13,5%
Bons (médios)	10	50% 68%
Fracos	4	20% 13,5%
Muito fracos	1	5% 2,5%
	<hr style="width: 50%; margin: 0 auto;"/> 20		<hr style="width: 50%; margin: 0 auto;"/> 100%	<hr style="width: 50%; margin: 0 auto;"/> 100%

Convém insistir em que o teimar em respeitar semelhante distribuição (*distribuição forçada*) não implica necessariamente um fracasso para um certo número de alunos. Todas as notas são *relativas* umas às outras. «Muito fraco», em certos grupos bons, pode situar-se acima da nota de reprovação. Em consequência, *é melhor adoptar letras de classificação* do que as fórmulas tradicionais.

2) *Lutar contra a contaminação e contra a tendência central*

Além dos perigos de estereotipia e dos efeitos de halo que descrevemos a p. 41, espreita ainda os avaliadores um outro fenómeno de contaminação, de natureza mais mecânica. Se utilizarem consecutivamente, para um mesmo aluno, uma série de escalas orientadas no mesmo sentido (por exemplo, do negativo ou do melhor ao mais fraco) e talvez mesmo apresentadas numa mesma página, as classificações tendem a agrupar-se dum mesmo lado.

É aconselhável utilizar para cada escala uma página diferente e tirar à sorte o sentido de apresentação de cada uma delas.

Finalmente, convirá deixar um intervalo bastante longo depois da avaliação de cada qualidade dum mesmo sujeito e

fazer avaliar uma mesma qualidade por tantos professores quanto possível.

Muitos avaliadores tem ainda tendência a agrupar as notas no centro da escala, com medo de exagerar no sentido positivo ou negativo ou com receio de suportar a responsabilidade dum fracasso. A regra da distribuição forçada presta neste caso úteis serviços. Mas para ser realmente eficaz, deve poder apoiar-se numa descrição precisa e variada dos diferentes escalões.

Também não se deve perder de vista que a distribuição gaussiana é pedagogicamente anormal (ver na quinta parte «O mito da curva de Gauss»).

Quando tiverem de ser combinadas as avaliações de muitos professores, os que se deixarem seduzir pela solução da facilidade que é a nota central devem saber que abdicam uma grande parte da sua influência em favor dos colegas que ousam utilizar o conjunto da escala.

F. Bacher nota com razão que a tendência central se manifesta sobretudo nas provas mais difíceis de classificar «que são talvez também aquelas em que as aptidões do candidato melhor se manifestam». Para nos convenceremos da pertinência dessa observação, basta comparar os resultados num ditado (em que uma tabela fixa de penalização conduz a todos os graus da escala) aos resultados atribuídos a um resumo de leitura. Guerbet-Seux e Reuchlin, citados por Bacher, observam um caso em que o peso real do ditado aumenta mais de 50% e em que o peso dos recontos diminui na mesma proporção, quando as duas notas se combinam na avaliação dum mesmo aluno.

F. Como sintetizar as avaliações


Ao fim dum período escolar, importa sintetizar as avaliações (que é desejável que sejam tão numerosas quanto possível), sem o que nem professores nem alunos saberão interpretá-las.

Voltar ao simples cálculo da média seria correr o risco de arruinar os esforços de rigor precedentes. Com efeito, avaliações ordinais, teoricamente pelo menos, não permitem nem adição nem subtração.

Eis um método simples que permite situar um aluno em cada disciplina, depois no conjunto das disciplinas.

No início dispõe-se do quadro seguinte:

Nota		Nível
A		I
B		II
C		III
D		IV
E		V



 Datas:

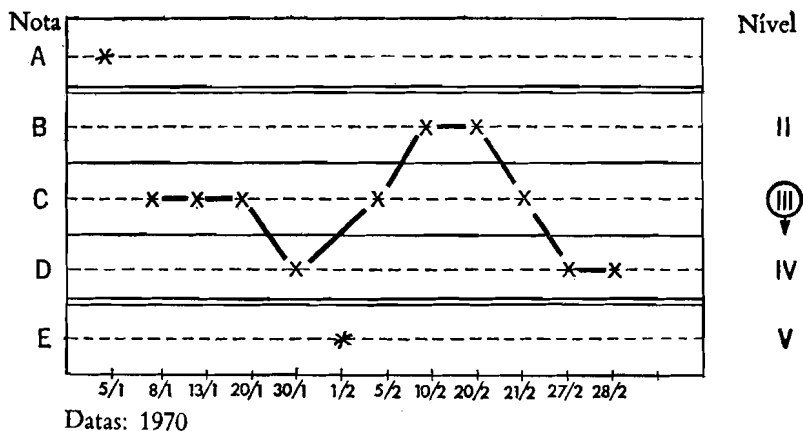
Um professor utilizará tantos quadros semelhantes quantos os aspectos que desejar distinguir no seu curso: ditado, gramática, ortografia, redacção...

À medida que vai efectuando as avaliações, faz uma cruz no quadro, indicando a data ao fundo.

No momento de fazer a síntese, começa-se por riscar (sem a apagar!) a cruz correspondente à nota mais alta e a que representa a nota mais baixa, a fim de eliminar notas que poderiam ser acidentais.

¹ F. BACHER, *La docimologie*, in M. REUHLIN, *Traité de psychologie appliquée*, 6, Paris, P. U. F., 1973, p. 33.

Suponhamos que se obtém o quadro seguinte



Interpretação

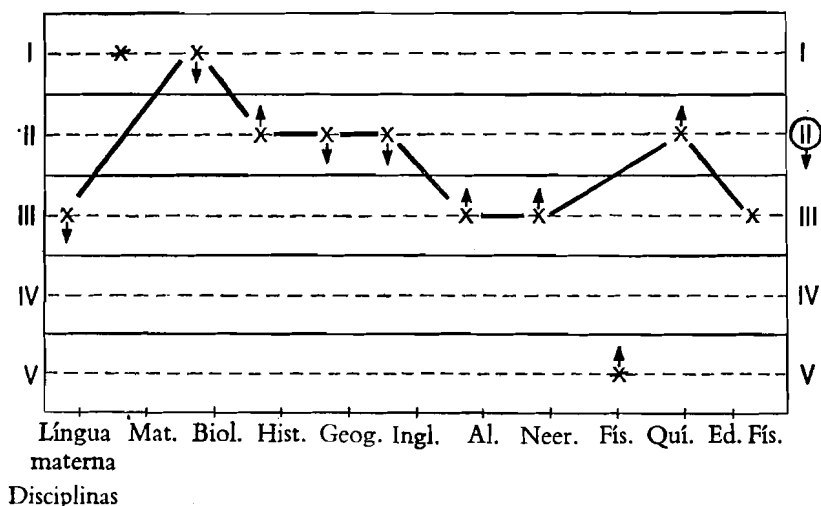
- 1) Os resultados concentram-se em três bandas: II, III e IV. Não são nem óptimos nem fracos ou insuficientes. A zona de concentração é marcada com traços duplos.
- 2) O maior número de notas (o modo) situa-se na banda III. É a tendência dominante.
- 3) Como a banda mais povoada, depois da III, é a IV, colocar-se-á uma flecha sobre a III, nessa direcção.
- 4) Reunindo as cruces, obtém-se um perfil que faz aparecer:
 - a) Uma razoável homogeneidade dos resultados. Se o aluno passeasse anarquicamente sobre as 5 bandas, teríamos de ver se é o único nesse caso (causa a determinar ao nível do aluno) ou se a maioria da classe se encontra nessa situação (ver se os trabalhos do professor não estão mal ajustados).
 - b) Uma descida no fim do período. Não é muito acentuada e pode explicar-se pela fadiga, por uma doença benigna,...

Como vemos, não basta verificar uma situação traduzida pelo perfil, é preciso descobrir o porquê. Deste modo, os comentários, orais e escritos, dos professores irão esclarecer e guiar tanto o aluno como os pais.

Note-se, contudo, que este sistema não pode funcionar se o professor se limitar a duas ou três avaliações. Ainda menos com uma só! Neste caso, será o comportamento de certos professores obnubilados pela quantidade de matéria a ensinar... ou alérgicos ao longo trabalho de classificação que a introdução duma nova forma de avaliação irá felizmente melhorar.

QUADRO PARA O CONJUNTO DAS DISCIPLINAS

PERÍODO DE A DE 19 ..



Interpretação

Encontram-se quatro notas de conjunto na banda II e quatro na banda III. É a tendência do perfil (aqui, para cima) que decidirá da dominante. Neste caso, II, com flecha para baixo.

Se houver igualdade prática entre dois níveis, pode-se ou decidir que o nível superior domine (sobretudo se se tratar de

dois níveis contíguos) ou cercar dois algarismos e reunir os dois círculos por um traço.

Observações

1.º Este sistema de síntese das avaliações não contém qualquer poder mágico! Se forem cometidos os erros indicados neste capítulo, a nota de síntese será provavelmente desprovida de sentido.

Um fenómeno de concentração nas bandas medianas surpreende muitas vezes os utilizadores. Esquecem que, no sistema tradicional, que leva a uma percentagem, as notas se concentram também na mesma zona, entre 65 e 75 ou 80%, permanecendo não raro quase vazio o resto da escala (de 100 escalões).

Na realidade, o fenómeno de concentração é muito mais surpreendente porque a escala de cinco graus tem uma visualização muito melhor que a antiga.

2.º Nada impede de ponderar as notas, isto é, por exemplo, dar duas vezes mais importância ao trabalho na aula do que aos trabalhos em casa, a determinado aspecto duma disciplina mais do que a outro, etc.

3.º *Na parte consagrada à «moderação», virão as indicações suplementares para explorar as notas, em particular para determinar os resultados e as classificações do fim do ano (ver pág. 236).*

G. Um caso particular:

A classificação da composição francesa

Este problema merece algumas considerações particulares, porque é dos mais difíceis.

Se estivesse resolvido, a docimologia obteria talvez a sua mais bela vitória. Infelizmente ou, melhor, felizmente para o

homem, a mais nobre das suas actividades, a avaliação da beleza, da verdade e do bem, escapará sempre à quantificação objectiva e, portanto, automatizável.

1) *Quatro métodos de avaliação*

Os principais métodos reduzem-se a quatro: método de impressão geral, escala dos espécimes, método analítico e contagens das frequências.

a) **O método da impressão geral**

É o mais subjectivo. É criticado pela falta de fidelidade dum mesmo examinador, pelos desacordos entre muitos examinadores, especialistas ou não, e pela sua ineficácia frequente. Muitos professores cansaram-se a classificar com cuidado centenas de trabalhos sem obterem resultados convincentes.

Muito haveria a dizer sobre a metodologia da composição francesa e sobretudo sobre a necessidade duma reforma profunda dos hábitos escolares que tratam cada aluno como um escritor em potência, mas não é esse o nosso objectivo. Insistamos simplesmente em que a conjugação de duas fraquezas muitas vezes acentuadas, uma docimológica e outra metodológica, carrega o exercício de redacção com uma hipotéca de tal modo pesada que há quem considere inútil esse exercício. Não somos dessa opinião.

Não obstante as deficiências, o método de impressão geral é o que está mais em harmonia com a complexidade essencial da composição. Utilizando uma escala reduzida, de 3 ou 5 graus no máximo, pode obter-se um notável acordo entre professores experimentados, com a mesma formação e pertencendo a um mesmo meio escolar. Veremos adiante as precauções a tomar antes de decidirmos duma reprovação.

Uma composição corrigida por este método, não deve, em nossa opinião, ambicionar melhorar muito o estilo ou a simples

correção da linguagem escrita. Há outros exercícios para esse efeito.

A composição tradicional deverá ser uma prova de excepção, destinada a descobrir talentos ou a controlar a maturação, bem como a identificar deficiências, falhas, defeitos que se combaterão com exercícios sistemáticos de enriquecimento e de correção.

Se decidirmos apenas pela apreciação global, merece atenção a *técnica de avaliação em tempo imposto*.

Essa técnica comporta variantes menores. O método proposto por Wiseman¹ é característico. Quatro professores avaliam independentemente trabalhos feitos em trinta minutos, à cadência de cerca de cinquenta exemplares hora (número imposto para os obrigar a uma decisão rápida). Calcula-se a média aritmética das notas.

Este método é provavelmente o mais eficaz para os trabalhos narrativos ou descritivos. Para as «dissertações», palavra que abrange todas as composições em que o sujeito deve argumentar, é mais discutível a superioridade do método. No entanto, parece ainda válido se os examinadores possuem alto nível intelectual.

A validade geral do método e a sua possibilidade de aplicação à «dissertação» é confirmada pelas dois estudos experimentais de J. Britton.

1.ª experiência²

Em primeiro lugar, foi classificado um conjunto de composições escritas por 168 alunos de 15 anos, segundo o método analítico, por um examinador experimentado. Em seguida, os mesmos trabalhos foram avaliados independentemente por oito professores, segundo o método da impressão geral em tempo imposto.

¹ S. WISEMAN, *The Marking of English Composition in Grammar School Selection*, in «British Journal of Educational Psychology», XIX, Nov. de 1949, 208.

² J. BRITTON, *Experimental Marking of English Composition Written by Fifteen-Year-Olds*, in «Educational Review (Birmingham)», vol. 16, I, 1963, 17-23.

Para dispor dum critério, foram classificados por dois examinadores cinco outros trabalhos de cada aluno.

As notas analíticas e a nota média de três impressões rápidas tiradas ao acaso de entre as oito de que se dispunha para cada estudante, foram então comparadas às notas de critério:

correlação entre notas analíticas e critério: .71

correlação entre notas rápidas e critérios: .76

Além disso, foi calculada a fidelidade da classificação rápida, comparando as notas de dois grupos de três examinadores, escolhidos ao acaso de entre os oito. Obtém-se um coeficiente muito elevado: .87.

2.^a experiência¹

Esta segunda experiência, muito mais perfeita que a precedente, incidiu sobre uma amostragem de 500 alunos de 17 anos, estratificada segundo o sexo, o tipo de escola secundária e o lugar de habitação (Londres, centro e bairros periféricos, grandes cidades, pequenas cidades). Eram apresentados também todos os temas de composição desde a simples narração à dissertação filosófica e à amplificação poética.

De novo se mostrou mais fiel e mais válida que a classificação analítica, a classificação múltipla segundo a impressão rápida.

b) Escalas de espécimes

Escolhe-se um pequeno número de composições, geralmente cinco, de forma a constituírem modelos representativos dos diferentes níveis de qualidade, desde o medíocre ao ótimo.

O professor avalia os trabalhos por comparação com os cinco textos-padrão da escala.

¹ J. BRITTON, N. MARTÍN et H. ROSEN, *Multiple Marking of English Composition*, Londres, H. M. S. O., 1966.

Depois de terem estado em moda em certos países¹, as escalas de espécimes foram abandonadas por várias razões².

1. É raro que as composições a classificar se assemelhem estreitamente aos textos-padrão.
2. Os examinadores têm tendência a só classificarem as características comuns que lhes interessam especialmente e desprezam as outras.
3. São necessárias escalas diferentes, de acordo com o tipo de assunto e com o nível dos alunos.

D. Pidgeon fez-nos ver como uma vasta experiência, feita recentemente na Inglaterra com uma escala de 5 graus, se concluiu com um desvio médio de 2 graus entre os examinadores. O resultado pouco melhor é do que o obtido em condições puramente subjectivas.

Note-se, contudo, que pelo menos a partir de 1903, o método das escalas por espécimes volta periodicamente às boas graças dos investigadores.

c) Método analítico

Dois ou três examinadores dão um certo número de pontos a cada uma das diferentes qualidades que se combina observar. Totaliza-se e calcula-se a média entre os examinadores.

Este método é o mais pesado dos quatro. É muitas vezes criticado, porque disseca um todo que precisamente parece escapar a toda a dissecação sistemática. S. Wiseman nota, aliás, que, num conjunto classificado analiticamente, o trabalho considerado unanimemente como o melhor pelos peritos, raramente figura em primeiro lugar.

¹ Para um estudo de conjunto, ver E. HINTON, *Study of the Qualities of Style and Rhetoric Found in English Composition*, New York, 1940.

² R. BRADOCK et al., *Research in Written Composition*, Champaign, III, N. C. T. E., 1973, p. 12.

Todavia, a investigação mostra que, se as qualidades a observar e a importância a conceder-lhes são definidas com precisão, o método analítico é o mais seguro, se tivermos que ficar-nos num só examinador. O método da impressão geral dá também bons resultados, como vimos, mas exige a participação de muitos examinadores ¹.

O método analítico, para ser eficaz e prático, parece dever obedecer a pelo menos duas exigências:

- 1.º Os examinadores devem aceitar o plano de análise. O ideal é que participam na sua elaboração.
- 2.º O plano não deve ser muito pormenorizado. Assim o o método apurado que utiliza E. Burton ² não tem interesse para a investigação, por causa do tempo extremamente longo que exige.

Que qualidades observar?

C. Remondino ³ demonstrou que as atitudes fundamentais dos avaliadores cultos (professores ou não) não diferiam essencialmente. A atenção deles incide sobre as mesmas qualidades. Mas atribuem-lhe, na apreciação de conjunto, valores diferentes. Nisso reside a origem das divergências.

Elaborar um quadro muito pormenorizado de análise do conteúdo é um engano, porque, embora os cambiantes distinguidos existam, os examinadores refundem-nos inconscientemente. A análise factorial põe esse processo em evidência. O trabalho de Remondino contém interessantes resultados a esse propósito.

Em primeiro lugar, Remondino interrogou longamente vinte professores, de disciplinas literárias (ensino secundário)

¹ Há também um bom estudo comparativo de B. CAST, *The Efficiency of Different Methods of Marking English Composition*, in «British Journal of Educational Psychology», IX, Nov. 1939, 257-259 e X, Fev. 1940, 49-50, citado por R. BRADDOCK, *o. c.*, p. 13.

² E. BUXTON, *An Experiment to Test the Effects of Writing Frequency...*, in «Alberta Journal of Educational Research», V, Junho, 1959, 91-99.

³ Estudo factorial sobre a classificação das composições escolares respeitantes à língua materna, in *Le Travail Humain*, XXII, Jan.-Junho, 1959, 27-40.

e elaborou também a lista das qualidades manifestadas em composições escolares. Concluiu pelo seguinte conjunto.

1. Legibilidade Qualidade duma escrita que se presta a uma leitura fácil, rápida, sem equívoco.
2. Estética Linha harmoniosa e agradável das letras e bom gosto na disposição em páginas.
3. Apresentação Limpeza, cuidado, boa apresentação
4. Exactidão da ortografia . . . Densidade dos erros de ortografia.
5. Exactidão morfológica . . . Densidade dos erros morfológicos.
6. Exactidão sintáctica Densidade dos erros de sintaxe.
7. Estrutura da exposição . . . Qualidade duma exposição feita com ordem, nas proporções desejadas e segundo um plano.
8. Riqueza de ideias Quantidade de ideias; recursos utilizados.
9. Pertinência das ideias Qualidade segundo a qual as ideias expostas estão em justa relação com o tema tratado .
10. Precisão da informação . . . Veracidade e exactidão objectiva das afirmações e dos factos expostos.
11. Exaustividade Qualidade que consiste em não deixar de lado nada do que deveria dizer-se.
12. Concisão Qualidade segundo a qual as coisas a dizer são apresentadas com o mínimo indispensável de termos, sem repetições, redundância ou tortuosidade.
13. Propriedade da linguagem . . Junto emprego dos termos.
14. Estilo Facilidade, exactidão, domínio da linguagem sob o ponto de vista da «construção das frases».
15. Originalidade Qualidade segundo a qual transparece e se afirma qualquer coisa da personalidade.
15. Maturidade Capacidade de julgar, profundidade crítica, agudeza dos raciocínios expostos.
17. Imaginação Capacidade de criação, de transfiguração, de «projectão» revelada no trabalho.

Em seguida, uma análise factorial mostrou que essas 17 qualidades diziam respeito a quatro grupos apenas:

1. Apresentação gráfica (1, 2 e 3);
2. Uso da língua (4, 5, 6, 13, 14);
3. Conteúdo e organização da exposição (7, 8, 9, 10, 11);
4. Aspectos pessoais do fundo (15, 16, 17).

Exemplos

As conclusões de Remondino confirmam largamente as dos outros investigadores. As escalas seguintes, construídas pela *Educational Testing Service* (E. T. S.) de Princeton¹, e ligeiramente retocadas por E. Page², testemunham a semelhança dos pontos de vista.

Traduzimos o texto completo porque nos fornece um dos melhores exemplos que actualmente se podem dar para indicar o caminho a seguir.

CRITÉRIOS PARA CLASSIFICAR AS COMPOSIÇÕES

I. Definição dos pontos a avaliar

- A. Ideias ou conteúdo — quantidade e qualidade do material utilizado para tratar do assunto.
- B. Organização — relação entre as partes e o todo da composição.
- C. Estilo — utilização da linguagem para além da simples correcção gramatical.
- D. Mecânica — ortografia, gramática, pontuação.
- E. Criatividade.

II. Guia para avaliação dos cinco pontos

- A. *Ideias ou conteúdo*

¹ E. T. S., *Definitions of Ratings on the E. T. S. Compositions Scale*, citado por E. PAGE, o. c., pp. 76-77.

² O. c., pp. 78-80.

Nível elevado

O estudante trata de todos os pontos indicados no tema ou plano de trabalho. Compreende bem o assunto e utiliza definições claras. Sabe considerar o assunto numa perspectiva mais ampla do que os outros alunos da classe. Por outras palavras, mostra uma experiência mais rica.

Nível médio

As ideias são apropriadas mas convencionais e pouco numerosas. Não são considerados certos aspectos do tema. O aluno não parece ter um espírito ricamente mobilado.

Nível baixo

O estudante omite muitos aspectos importantes do tema. Parece não dispor duma reserva de conhecimentos relativos ao assunto e, em consequência, repete continuamente algumas ideias simples.

B. Organização

Nível elevado

O estudante segue um plano definido. Se apresenta o pró e o contra, aduz razões pertinentes, numa ordem eficaz. Se descreve qualquer coisa, fá-lo de maneira ordenada (do cimo para a base, por ordem de importância, por ordem de complexidade, etc.). Se o estudante explica um conceito ou um processo, utiliza um plano coerente de análise, de definição ou de ilustração. O estudante tem a noção do que se relaciona com o seu plano e evita repetições. Mostra sentido de equilíbrio ao tratar as diferentes partes do trabalho.

Nível médio

O estudante não se cinge ao seu plano ou mistura ideias sem relação com o assunto. Consagra muito tempo a coisas pouco importantes ou repete-se. Trata o tema por associação livre (o que é que me vem à mente quando penso em Hawai?) de preferência a seguir um objectivo bem definido.

Nível baixo

O estudante não parece ter-se perguntado o que iria fazer, antes de começar a escrever. Não segue um plano. O trabalho toma uma direcção, depois muda, muda outra vez, outra ainda, até que o leitor se perde. Os pontos principais não são claramente separados uns dos outros e a sua ordem de apresentação fica ao acaso.

C. *Estilo*

(Podem intervir, na avaliação, muitos aspectos do estilo: individualidade, vivacidade, elegância, etc. Todavia, interessamo-nos aqui em três aspectos estilísticos apenas: clareza, variedade, e soma de recursos linguísticos).

Nível elevado

O estudante utiliza uma linguagem que torna fácil a compreensão do trabalho. Utiliza palavras adequadas no seu sentido habitual. As palavras são apresentadas em ordem normal. As transições são bem feitas. O aluno evita as ambiguidades e não ilude a expectativa do leitor. Ao mesmo tempo, evita a repetição monótona de palavras, de complementos ou de estruturas de frases. Finalmente, revela o conhecimento duma ampla gama de recursos linguísticos. O vocabulário é bom. Utiliza estruturas paralelas e faz um uso subtil da subordinação.

Nível médio

O estudante desorienta às vezes o leitor, utilizando uma palavra inadequada ou uma construção bizarra, ou então utilizando uma metáfora pouco clara, ou deslocando de maneira inoportuna um complemento ou uma subordinada, ou ainda fazendo transições abruptas. A repetição de palavras, de maneiras de dizer e de estruturas de frases, torna-se monótona. Os recursos linguísticos são limitados. O aluno utiliza de preferência clichés e expressões gastas.

Nível baixo

As palavras são utilizadas de maneira vaga. Construções ambíguas, claudicantes, vocabulário e estruturas de frases infantis.

D. *Mecânica*

Nível elevado

A estrutura das frases é habitualmente correcta, mesmo quando se trata de modelos variados e complicados. Respeitam-se as regras de ortografia, e as próprias palavras difíceis são geralmente escritas sem erro. Não há violação grave das regras de pontuação, de maiúsculas, de abreviaturas, de grafia dos números.

Nível médio

Erros de sintaxe ocasionais. As palavras difíceis são geralmente mal ortografadas. Algumas violações de regras de pontuação, etc.

Nível baixo

Grande número de erros.

E. *Criatividade*

Nível elevado

O estudante surpreende pelas maneiras novas e eficazes de considerar o problema. Introduce ideias novas ao tratar do assunto. Descobre maneiras originais e interessantes de utilizar a linguagem para pôr em realce as ideias.

Nível médio

O estudante pensa naquilo que esperariam que ele pensasse. Trata das coisas como toda a gente mais ou menos. Utiliza expressões e estruturas de frases ordinárias.

Nível baixo

O estudante utiliza clichés de pensamento e de expressão. Trata o tema de maneira superficial. Repete fórmulas sem lhes compreender realmente o significado.

Um grupo de professores que decidisse avaliar uma mesma composição, tendo em conta as cinco escalas que acabamos de referir, obteria como resultado notas ainda mais discordantes. Porquê? Porque os limites entre os diferentes graus permanecem indistintos (não consideramos neste caso o problema da síntese das notas relativas às cinco escalas).

Para atingir um grau de concordância elevado, é necessária ainda mais precisão. Demos um exemplo¹.

À guisa de composição, convidaram-se alunos de doze anos a preencher uma lacuna numa narração de aventuras da qual se dera o resumo dos capítulos precedentes e depois duas páginas completas entre as quais se deixara uma página em branco.

¹ G. DE LANDSHERE, R. DZ BAL, J. PAQUAY-BEKKERS, *Construction d'échelles d'évaluation pour l'enseignement secondaire rénové*. Laboratoire de Pédagogie expérimentale de l'Université de Liège (investigação em curso).

Avaliou-se um mesmo trabalho segundo cinco dimensões:

- Pertinência semântica da resposta
- Pertinência sintática
- O vocabulário
- As estruturas gramaticais
- A organização das ideias

A pertinência semântica foi considerada da seguinte maneira. Note-se que a abordagem foi de algum modo programada. Não basta um modo de emprego correcto: todos devem compreendê-lo e aplicá-lo da mesma maneira. Ora, no decurso da experiência, constatou-se que subsistiam diferenças de avaliação, não já por causa da imprecisão das indicações mas em virtude da sua má compreensão.

PERTINÊNCIA SEMÂNTICA

A. DEFINIÇÃO DA DIMENSÃO

O trabalho do aluno é pertinente sob o ponto de vista semântico se desenvolve ideias apropriadas ao contexto, isto é, o resumo e os parágrafos 1 e 3. Não deve necessariamente conter todas as informações do texto do autor. O importante é que não os contradiga.

Duas perguntas permitirão avaliar a pertinência semântica:

- O aluno não utilizou elementos que contradigam as informações?
- Integra-as num contexto mais vasto?

Descrevem-se em seguida estes dois aspectos.

B. DEFINIÇÃO DOS ASPECTOS QUE INTERVÊM NA DIMENSÃO

1.º aspecto: *Introdução de elementos não pertinentes*

Os elementos não pertinentes são elementos que entram em **contradição** com as informações dadas pelo texto.

A leitura de um número significativo de exercícios de alunos permitiu-nos estabelecer a lista destes elementos não pertinentes. Servirá para avaliar este primeiro aspecto da pertinência semântica.

1. O início do parágrafo 2 não situa os factos no laboratório (Exemplo: empurrar a porta, entrar... são acções não pertinentes. Já estão expressas no parágrafo 1).
2. Não se faz qualquer alusão ao «espectáculo insólito». Essa alusão pode limitar-se a uma exclamação, a uma única palavra.

3. No parágrafo 2, está **presente** ou é esperado o comissário.
4. São modificados o aspecto e o conteúdo da mensagem (Por exemplo, o bocado de papel torna-se uma carta, longa e cuidada).
5. A criança evoca ferimentos graves, sinistros, catástrofes...

2.º aspecto: *Integração das informações*

Não basta a simples avaliação do número de elementos **não** pertinentes para dar conta do trabalho do aluno.

Afinar-se-á esta avaliação, apreciando a maneira como o aluno integra os factos que relata num contexto mais vasto.

- a) Iremos encontrar os índices desta integração
 - na evocação da personalidade do tio;
 - nas alusões a personagens secundárias;
 - na expressão dos sentimentos de Lucas...
 (*Esta lista não é exaustiva*).
- b) Considerar-se-á que o aluno não integrou as informações se se limitou a relatar estritamente factos que se situam depois da entrada de Lucas no laboratório (fim do primeiro parágrafo) e a sua resposta ao comissário (princípio do terceiro parágrafo).
- c) Alguns trabalhos desenvolvem ideias que, sem estarem necessariamente em contradição com os parágrafos 1 e 3, não têm relação directa com a situação.

C. DESCRIÇÃO DA ESCALA: *Pertinência semântica*

	<i>Introdução de elementos não pertinentes</i>	<i>Integração das informações</i>
MB	Sem elementos não pertinentes	Integração dos factos do § 2 num contexto mais vasto
B	Sem elementos não pertinentes	Sem integração
	1 ou 2 elementos não pertinentes	Integração dos factos do § 2 num contexto mais vasto
S	1 ou 2 elementos não pertinentes	Sem integração
F	3 ou 4 elementos não pertinentes	Quaisquer que sejam as características de integração
I	Mais de 4 elementos não pertinentes	Quaisquer que sejam as características de integração
	Qualquer que seja o número de elementos não pertinentes	Ideias sem relação directa com a situação

D. TENTATIVA DE AVALIAÇÃO DA DIMENSÃO

1. Faça incidir a apreciação sobre o primeiro aspecto:

- sem elementos não pertinentes?
- 1 ou 2 elementos não pertinentes?
- 3 ou 4 elementos não pertinentes?

2. Faça incidir a apreciação sobre o segundo aspecto:

- integra as informações num contexto mais vasto?
- não integra as informações?
- desenvolve ideias sem relação com a situação

3. Tendo em conta o esquema seguinte, escolha o nível que corresponde às duas características que identificam no trabalho do aluno.

<i>Integração de elementos não pertinentes</i>	<i>Integração</i>
Sem elementos não pertinentes	→ integração → MB
	→ sem integração → B
	→ ideias sem relação directa com a situação → I
1 ou 2 elementos não pertinentes	→ integração → B
	→ sem integração → S
	→ ideias sem relação directa com a situação → I
3 ou 4 elementos não pertinentes	→ ideias sem relação directa com a situação → I
	→ ideias sem relação directa com a situação → I
Mais de 4 elementos não pertinentes	→ ideias sem relação directa com a situação → I

Nota para a PERTINÊNCIA SEMÂNTICA:

d) Método das contagens de frequências

Indicamo-lo aqui, embora não tenha lugar na categoria das avaliações subjectivas, para completar o quadro dos métodos.

Consiste em definir tipos de erros (após um estudo diagnóstico), em contar estes na composição, exprimindo o número obtido em 100 ou 1000 palavras.

Em vez de adoptar um ponto de vista negativo, podemos também fazer o inventário (relativamente exaustivo) dos diversos processos utilizados na composição. O melhor exemplo de tal método é dado nos *Eléments pour servir à l'étude d'analyses littéraires* de A. Purves¹.

Um último método positivo consiste em isolar certos aspectos que se revelam anunciadores, a um nível dado, da nota normalmente atribuída por examinadores qualificados. Pode então encarregar-se da classificação um ordenador programado para esse efeito. Foi o que fez E. Page.

As contagens de frequências oferecem provavelmente a única esperança de classificação objectiva das composições. Parece, no entanto, que devem permanecer no domínio da investigação estranho à prática escolar, a não ser que o ordenador se torne de acesso de tal modo fácil que o estudo das frequências se faça de maneira económica e constitua um dos aspectos da avaliação. O que de forma alguma se considera excluído.

2) *Vários temas à escolha?*

J. Britton nota com razão que a redacção de exame sobre um tema único implica muitas vezes da parte do professor o admitir a hipótese de que um tema como «Nuvens» revelará tão bem a aptidão dum estudante como «Exponha as suas ideias sobre a pena capital».

Demonstrou-se experimentalmente que não era verdade. Voltaremos ao assunto. Em todo o caso, é duvidoso que a justiça seja muito mais respeitada se se oferecerem três temas à escolha quando os pontos de exame são atribuídos por uma redacção única. Seria preciso provar primeiramente que com três e mesmo cinco temas se abrange todo o espectro dos estímulos necessários para dar a cada um uma igual probabilidade

¹ A. PURVES e V. RIPPERE, *Eléments pour servir à l'étude d'analyses littéraires* (traduzido do inglês por J. Dubois), in «*Scientia Paedagogica Experimentalis*», VI, 2, 1969, 292-334.

E mesmo que se conseguisse, importaria ainda que o professor seleccionasse sempre os temas com perfeito conhecimento de causa.

Como vínhamos dizendo, a experiência mostra que o rendimento do aluno varia com o tema proposto. Finlayson¹ estabeleceu, para os últimos tempos da escola primária, que se os alunos forem convidados a fazer uma redacção a partir duma mesma escolha de quatro temas com intervalo duma semana, os que mudarem de tema (mais de metade na experiência) obtêm notas significativamente diferentes.

Wiseman e Wrigley² chegam à mesma conclusão (escolha de entre o mesmo conjunto de assuntos a quatro meses de intervalo), provando que a diferença de nota não é imputável à falta de fidelidade do professor. Vernon e Mallican³ confirmaram a conclusão ao nível universitário.

Que fazer? Se quisermos verificar a capacidade de expressão por escrito numa perspectiva estreitamente definida (por exemplo: capacidade de fazer uma reportagem científica ou de escrever uma carta comercial) pouco importará a sensibilidade particular dos candidatos e um tema basta.

Ao contrário, se se tratar ou de revelar uma aptidão para a expressão onde quer que ela exista ou de levar os alunos a discutir problemas em que a feição de espírito e o nível de informação podem desempenhar um grande papel, é quase impossível fazer um juízo a partir duma única composição. O ideal será, pois, ter em conta também o trabalho do ano. Se de cada vez o aluno tiver podido escolher de entre vários temas (cinco, por exemplo), terá reagido a várias

¹ D. S. TINLAYSON, *The Reliability of the Marking of Essays*, in «British Journal of Educational Psychology», XXI, 2, 1951, citado por Britton (1966).

² S. WISZKAN e J. WRIGLEY, *Essay-Reliability: the Effect of Choice of Essay-Title*, in «Educational and Psychological Measurement», 18, 1, 1958.

³ P. VERNON e G. MILLIGAN, «A Further Study of the Reliability of English Essays», in «British Journal of Statistical Psychology», VII, 2, 1954.

dezenas de temas diferentes no decurso de alguns meses e pode supor-se que todos tenham tido a sua oportunidade. Permitir a livre escolha do tema a tratar seria tentador, se fosse possível prevenirmo-nos contra a preparação fraudulenta. É bem difícil...

3) *Conclusão*

A sorte a reservar à composição francesa dependerá do objectivo em vista. O que é que se pretende? Corrigir um defeito particular? Nesse caso, é melhor examiná-lo apenas a ele nos exercícios ou atacá-lo com trabalhos especiais.

Conhecer a capacidade que os alunos têm de se corrigirem, de identificarem a prosa de acordo com o uso correcto? Nesse caso, um teste objectivo com respostas de escolha múltipla, pode muito bem servir.

Fazer adquirir a capacidade de se exprimirem claramente? Nesse caso, temas muito limitados permitem concentrar a atenção no objecto do trabalho. Porque banir assuntos tais como: «Descreva uma bicicleta de corrida»?

Esquece-se muitas vezes uma indicação crucial: quem será o leitor? Na realidade, o estilo varia, de maneira mais ou menos considerável, segundo o destinatário da mensagem. Porque razão não dar sempre esta indicação ao propor um tema de composição?

Revelar talentos de escritores? Será preciso dar livre curso à expressão do aluno, convidando-o muitas vezes a escolher ele o assunto.

Mas não nos iludamos. Dessa maneira iremos fazer o teste a capacidades diferentes. Na segunda experiência de Britton, descrita acima, o autor demonstrou:

1.º que a correlação entre os resultados na redacção criativa e um resumo de texto varia de .30 a .40;

2.º que a correlação da redacção com um teste de compreensão de texto é um pouco mais elevada, de .35 a .45.

Os desvios são, pois, consideráveis.

Enfim, J. Foley¹ nota que todos parecem convencidos intuitivamente de que a melhor maneira de avaliar a capacidade de escrever é fazer escrever. Ora, nós dispomos de provas cada vez mais numerosas de «que um teste de aptidão verbal permitiria prever as notas de composição melhor do que um teste que exigisse do estudante que escrevesse realmente».

Já dissemos que não pretendíamos banir a composição tradicional. Em certos países onde fora suprimida por causa da dificuldade de avaliação rigorosa, tornou a ser introduzida, designadamente nos Estados Unidos em que a Comissão dos Exames de Admissão ao Ensino Superior renunciou ao «Teste objectivo de composição».

Todavia, se a redacção continua a desempenhar um papel importante nos exames, a maneira ainda muito difundida de classificar é que deve ser modificada. Fazemos nossa a afirmação de Braddock:

«Se tivermos de avaliar um grande número de estudantes para decidir do bom ou mau resultado, importa dar ocasião aos que reprovam de fazer uma segunda prova (...). Se pretendermos atribuir notas, serão necessários três trabalhos pelo menos. Escolher-se-ão os dois melhores e far-se-á a média (...). Os trabalhos devem ser classificados por dois examinadores pelo menos, utilizando um sistema bem compreendido e aceite, no qual foram bem treinados»².

¹ J. FOLEY, in B. BLOOM, J. HASTINGS et G. MADAUS, *Handbook...*, o. c., p. 800-801.

² R. BRADDOCK et al., o. c., p. 45.

III. Classificação objectiva

A classificação objectiva não faz intervir a opinião pessoal dos examinadores. O exemplo mais simples é-nos dado por uma pergunta contendo dez multiplicações à razão dum ponto por resultado exacto. Do mesmo modo, a classificação dum teste composto de *items* de escolha múltipla opera-se por simples contagem do número de escolhas correctas.

De modo mais geral, chama-se classificação objectiva «a atribuição de valores numéricos a amostragens comportamentais suficientemente limitadas, definidas e controladas para permitir um acordo geral entre os avaliadores ou classificadores».

Como notámos acima, a objectividade da classificação não garante a objectividade do exame no seu conjunto.

A escolha das perguntas, em última análise, continua sempre subjectiva e facilmente se imagina um exame no qual nenhuma das perguntas satisfaça a uma amostragem real da capacidade que se pretende avaliar. Por mais objectiva que possa ser a correcção, a prova não teria, neste caso, a mínima validade. Abundam os exemplos de semelhante dificuldade.

IV — A aferição ou medida da posição relativa

A interpretação de uma nota seja ela qual for exige um ponto de referência, um critério, uma norma.

Aferir consiste em definir normas.

No decurso dos últimos cinquenta anos, considerou-se como norma sobretudo a distribuição das *performances* individuais em relação a uma tarefa ou conjunto de tarefas. A partir do *score* obtido, verifica-se que um estudante está situado no 70.º lugar, por exemplo, numa população de referência cuja distribuição de resultados se opera no total de cem. Por conseguinte, este 70.º lugar não indica o que o aluno em causa conhece da matéria do teste (conheceria talvez oito décimos),

mas antes o lugar que ocupa numa corrida de cem participantes que pertencem à sua categoria (por exemplo, os alunos da região de Paris que frequentam o CM2). O critério de apreciação, a norma, é, neste contexto, ou a posição na ordem de classificação, ou, em processos mais rigorosos, a situação em relação à média do grupo em referência. Os testes assim aferidos têm vindo a ser chamados cada vez com mais frequência *testes normativos* (*norm referenced tests*), designação que não é efectivamente a mais feliz mas que já entrou no uso corrente.

Há uns anos que apareceu uma nova tendência que cada vez se afirma mais na prática escolar.

A primeira preocupação da educação de base não é competitiva ou selectiva, mas formativa. Para ajudar um aluno, importa bem menos indicar-lhe o lugar que ele ocupa em relação aos outros, na corrida para o saber do que ensinar-lhe até que ponto os seus esforços conduzem ao caminho que leva ao domínio desejado duma aprendizagem. A que distância se encontra ainda do objectivo a atingir e que obstáculos deverá ainda ultrapassar?

Nesta perspectiva, o critério, a norma, é a tarefa indicada e a aferição far-se-á segundo as componentes dominadas. Os testes destinados a fornecer este tipo de informação testes centrados sobre os objectivos (*criterion referenced test*,)¹.

¹ «Eu chamo medidas de referência criterial aquelas que se relacionam com um valor absoluto de qualidade, ao passo que medidas de referência normativas serão as que se relacionam com um valor *standard* relativo».

IV-1. Aferição dos testes normativos

Como acabamos de ver, a norma é estabelecida aqui essencialmente mediante as *performances* dum grupo de indivíduos.

Segundo o fim em vista, as dimensões desse grupo variam: todos os alunos ou uma amostragem dos alunos do mesmo nível escolar ou da mesma idade, numa cidade, num concelho, numa região, num país... Rigorosamente, a validade das normas limita-se aos grupos que lhes deram origem. Importa, portanto, saber quais.

Já vimos que uma das maneiras de proceder consiste em informar o aluno do lugar que ocuparia se o seu grupo contasse cem indivíduos (centilagem). Por razões a que voltaremos, prefere-se muitas vezes situar uma *performance* em relação à média ou mediana¹ dum grupo em referência.

A partir desses pontos de referência, constroem-se escalas que permitem exprimir *scores* brutos em unidades comparáveis.

Os sistemas de aferição normativos mais conhecidos e mais acessíveis são os seguintes:

1. Centilagem;
2. Notas *standards* ou notas Z;
3. Escala normalizada de cinco classes;
4. Escala normalizada de 9 classes (Stanines).

Vamos estudá-los sucessivamente.

A. Centilagem

1) Definição

A centilagem, ou posição ocupada numa escala de cem graus, tem ainda uma certa aceitação em virtude da semelhança

² Recordemos que a mediana é o ponto que divide uma série de notas ordenadas em duas partes iguais, 1-2-3-4-5 ou 1-2-3-.4-5-6.

exterior com a percentagem, da qual todavia difere de modo fundamental. A percentagem indica que proporção do total dos pontos atribuídos num exame o aluno obteve. O centil indica quantos alunos se classificariam após um aluno dado se a classe tivesse cem alunos. Portanto um *score* equivalente ao 90.º centil é superior a 90% da proporção considerada.

A mediana corresponde ao 50.º centil.

2) Cálculo

Em princípio, não se calculam centis com menos de cem notas. O exemplo seguinte utiliza as cem notas que já nos serviram na altura do estudo da curva de Gauss.

Técnica geral

- Construir o quadro das classes (começando pela inferior);
- Indicar as frequências (f);
- Calcular as frequências acumuladas (f_c): cada frequência é somada à anterior.
- A partir destes dados, podemos calcular os centis com o auxílio da fórmula seguinte ¹:

$$C_x = \frac{1 + (N. C)/100 - f_c}{f_s} \times i$$

¹ $(N. C)/100$ = valor da nota (x) em centis.

C = posição da nota correspondente a um centil qualquer.

Exemplo: o 50.º centil corresponde à nota mediana, ou seja $N/2$. Para os outros, trata-se duma simples regra de três. O 75.º centil será portanto: $(100 \times 75) : 100 = 75$ (os números são os mesmos porque, no nosso exemplo, há precisamente 100 notas).

l = a média entre o limite superior da classe em que se encontra o centil procurado e o limite inferior da classe que o precede.

f_c = frequência acumulada da classe que precede aquela em que se encontra o centil procurado.

f_s = a frequência simples do intervalo em que se encontra o centil procurado.

i = intervalo de classe.

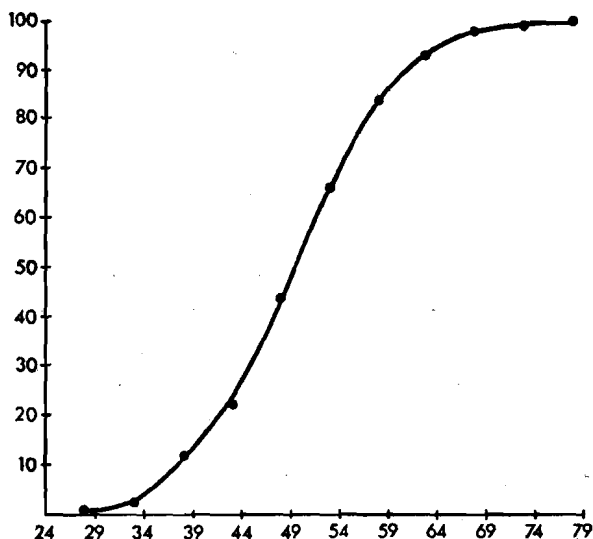
Todavia, na prática, contentamo-nos muitas vezes com valores aproximados que facilmente se lêem num abaque construído da maneira seguinte em papel milimétrico:

- 1.º Apresentar horizontalmente o valor das classes;
- 2.º Marcar com um ponto as frequências acumuladas nos limites superiores exactos das classes;
- 3.º Ligar os pontos por uma curva que, se os *scores* se distribuírem dum maneira bastante normal, toma a forma dum S (ogiva de Galton). Se a ogiva ficar bem traçada, podemos permitir-nos regularizar, «polir» o seu traçado;
- 4.º Basta nesse caso partir da escala vertical para ir ler, a partir do ponto correspondente na ogiva, o valor aproximado do centil procurado.

Exemplo:

<i>Classes</i>	<i>Frequências (f)</i>	<i>Frequências acumuladas (f_c)</i>
74 - 78	1	100
69 - 73	1	99
64 - 68	5	98
59 - 63	9	93
54 - 58	18	84
49 - 53	22	66
44 - 48	23	44
39 - 43	10	21
34 - 38	8	11
29 - 33	2	3
24 - 28	1	1

Abaco.



Comparação de alguns valores obtidos:

	<i>por leitura sobre o abaco</i> ¹	<i>por cálculo</i>
16.º centil	40,5	41
50.º centil	49,5	49,9
84.º centil	58,5	58,5

Observações:

1. O 50.º centil corresponde à mediana.
2. As vezes é preferível aferir em decis: o 10.º centil = = 1.º decil, etc.
3. Desvio interquartil = 75º C-25º C.
4. Valor aproximado do desvio-padrão = 3/4 do desvio interquartil.

¹ Recordemos que, na prática, a ogiva é desenhada sobre papel milimétrico. Quanto maior é o desenho tanto mais fácil e precisa é a leitura.

3) Crítica

Como já dissemos, a escala em centis não é mais que uma extensa escala de avaliação com 100 graus. Informa sobre o lugar ocupado mas não sobre a distância entre lugares. Ora, se a distribuição for normal, essa distância encurta à medida que nos aproximamos da mediana (igual à média no caso de normalidade perfeita).

O quadro seguinte mostra claramente o fenómeno.

	-2σ	-1σ	0	$+2\sigma$	$+2\sigma$
Centis	2°	16°	50°	84°	98°

Verifica-se que se aglomeram 68 centis à volta da média, o que dá uma classificação muito grosseira para a maioria do grupo. Corremos o risco de considerar, por exemplo, que o 75.º centil é separado do 25.º por uma diferença importante. Subsiste tenazmente nos espíritos a antiga norma da percentagem e leva a pensar que o 75º C equivale a três vezes o 25º. O quadro que vem a seguir mostra que efectivamente se trata de dois resultados relativamente próximos e médios.

Com efeito, se as normas em centis continuam a ser utilizadas, designadamente para a selecção em certas universidades americanas, é porque estas só recrutam na zona dos dez ou vinte centis superiores, zona em que a discriminação é satisfatória.

B. As notas standard ou notas Z

Um aluno obteve as notas brutas seguintes¹. Cálculo: 22 em 25; Leitura: 72 em 100; Ciências: 26 em 50.

O exame destas notas não nos diz grande coisa. Se o professor é severo na leitura e muito menos em cálculo, 72/100 em leitura é talvez mais meritório que 22/25 em cálculo.

¹ Adaptado de R. THOMAS, *o. c.*

Além disso, as notas brutas não permitem comparação com outros alunos do mesmo nível pedagógico, por exemplo.

Para tornar possível a comparação, exprimem-se essas notas em função de desvios-padrão, o que permite situá-las na mesma curva (notas aferidas ou *scores standard*).

Suponhamos que se obtêm os seguintes resultados:

	Média	Desvio-padrão
Cálculo	15	2,5
Leitura	50	10
Ciências	29	5

Cálculo dos *scores standard*: $z = (x - M)/\sigma$.

Exemplo:

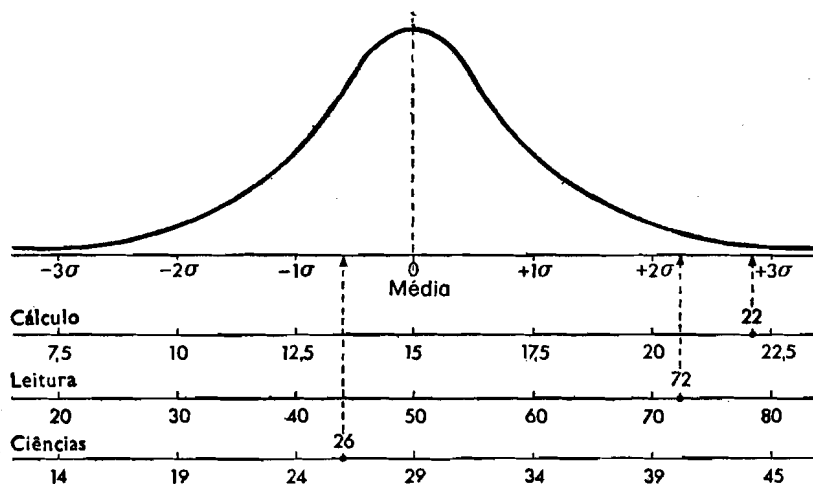
$$(22 - 15): 2,5 = 2,8 \sigma.$$

Conhecendo z , podemos encontrar a nota bruta:

$$x = M + z\sigma$$

$$x = 15 + (2,8 \times 2,5) = 22$$

Representação gráfica:



Aplicação:

O estudantes I e II foram submetidos a uma série de testes. Comparar os resultados ¹:

TESTE	Média M	Desvio padrão σ	X Scores brutos		x Desvios (M-X)		Scores standard z	
			I	II	I	II	I	II
Inglês	155,7	26,4	195	162	+ 39,3	+ 6,3	+ 1,49	+ 0,24
Leitura	33,7	8,2	20	54	- 13,7	+ 20,3	- 1,67	+ 2,48
Informação	54,5	9,2	39	72	- 15,5	+ 17,5	- 1,67	+ 1,88
Aptidões escolares	87,1	25,8	139	84	+ 51,9	- 3,1	+ 2,01	- 0,12
Atenção	24,8	6,8	41	25	+ 16,2	+ 0,2	+ 2,38	+ 0,03
Totais médias			434	397			+ 2,54 + 0,51	+ 4,51 + 0,90

Se considerássemos simplesmente os totais (scores brutos I e II), o estudante I seria considerado como superior ao estudante II. O cálculo das percentagens confirmaria essa impressão.

Ora se, mediante o cálculo dos *scores Z*, se reduzem todos os resultados à mesma média e à mesma unidade de desvio em relação a esta, verifica-se:

- 1.º que em média o estudante II (0,90) obtém um resultado quase duas vezes tão favorável como o estudante I (0,51);
- 2.º que o estudante II obtém os resultados mais homogêneos.

Crítica

As transformações em *scores Z* assenta na hipótese segundo a qual, para o grupo considerado, as aptidões que os testes

¹ Números extraídos de J. P. GUILFORD, *Fundamental Statistics*, o. c., p. 513.

definem se distribuem do mesmo modo e conduzem às mesmas médias e dispersões. Esta hipótese é praticamente impossível de verificar, mas, como nota J. P. Guilford, o certo é que esse sistema se mostra de qualquer modo melhor do que o cálculo da percentagem.

Além disso, o centrar da escala sobre a média dá uma imagem discriminativa falsa. Em virtude das imperfeições inevitáveis da classificação, o puro acaso fará oscilar um número mais ou menos elevado de alunos para um lado ou para outro.

Uma transformação

O scores Z tomam zero como média e um desvio-padrão igual a 1 como unidade.

Alguns consideram este sistema pouco sugestivo e preferem exprimir os resultados à maneira dos quocientes intelectuais em que 100 é tradicionalmente tomado como média e 145 considerado como um quociente de «génio».

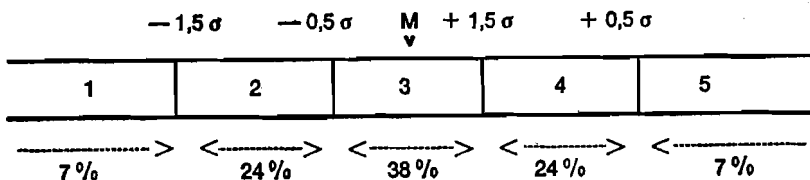
A transformação é fácil. Cem é tomado pois como média e cada score Z multiplicado por 15.

C. Escala de 5 classes (N entre 50 e 100)

As 2.^a, 3.^a e 4.^a classes têm um desvio padrão.

A 3.^a classe está centrada sobre a média.

As 1.^a e 5.^a classes têm extensão ilimitada.



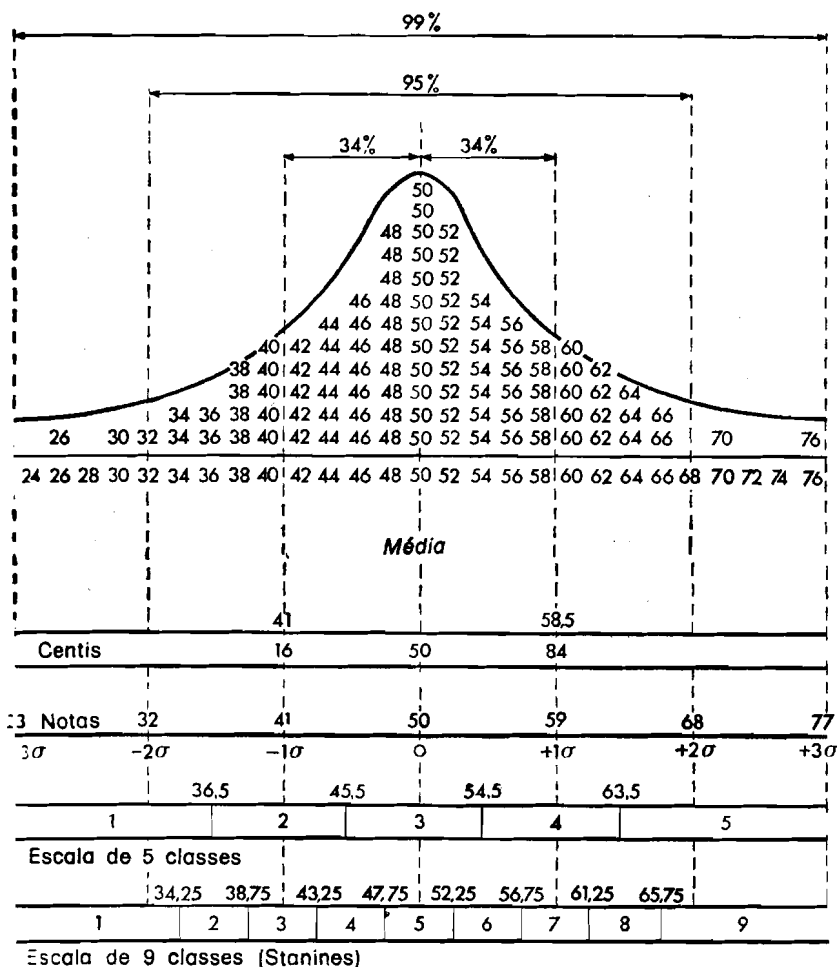
D. Escala de nove classes (Stanines) ($N > 100$)

As 1.^a e 9.^a classes são de extensão ilimitada.

As outras classes apresentam um meio desvio-padrão; a 5.^a está centrada sobre a média.

M								
$-1,75\sigma$	$-1,25\sigma$	$-0,75\sigma$	$-0,25\sigma$	0	$+0,25\sigma$	$+0,75\sigma$	$+1,25\sigma$	$+1,75\sigma$
1	2	3	4	5	6	7	8	9
\longleftarrow	\longleftarrow	\longleftarrow	\longleftarrow	\longleftarrow	\longleftarrow	\longleftarrow	\longleftarrow	\longleftarrow
4%	7%	12%	17%	20%	17%	12%	7%	4%

O gráfico que segue permite uma comparação fácil dos quatro tipos de normas.



IV-2. A aferição em relação ao objectivo

Passando da aferição tradicional à aferição centrada sobre os objectivos, abandona-se um critério externo (as *performances* do grupo de referência) por um critério interno (grau de domínio duma aprendizagem).

Nesta segunda perspectiva, o critério identifica-se ao objectivo pretendido, quer se trate de macro-objectivo (saber ler) ou de micro-objectivo (poder identificar as palavras que começam pela mesma letra). Entre estes dois extremos, distinguem-se também objectivos intermediários ou instrumentais.

Ao nível dos macro-objectivos, o problema da avaliação varia consideravelmente conforme se trate duma aprendizagem de algum modo fechada (dominar *toda* a técnica da leitura, objectivo convergente que implica um número de micro-objectivos e, portanto, de *items* acabados), ou duma aprendizagem aberta (aprender a exprimir-se criativamente pelo desenho, objectivo divergente).

Neste segundo caso, só podemos considerar os objectivos instrumentais, isto é, ver se o indivíduo dominou as técnicas, os instrumentos necessários à expressão criadora. Já vimos que as produções criadoras estão sujeitas em última análise a apreciações subjectivas.

O processo adoptado por A. Inizan¹ para decidir se um aluno sabe ler oferece um belo exemplo de macro-avaliação. Sabe ler, propõe ele, todo o indivíduo capaz de decifrar o texto-critério que figura no meu livro².

Facilmente se imagina um certo número de casos semelhantes em que o comportamento final pode ser avaliado com a ajuda dum critério preciso. Para os outros, os esforços de objectivação deverão muitas vezes centrar-se em objectivos intermediários.

¹ A. INIZAN, *Le temps d'apprendre à lire*. Paris, Bourrelier, 1964.

² Não se discute aqui a validade do critério proposto. Neste ponto só nos interessa a técnica geral.

Actualmente, o interesse está mais centrado nos micro-objectivos porque parecem permitir um progresso considerável da prática escolar.

Como vimos no capítulo consagrado ao encaminhamento objectivo para a formulação das perguntas, tende-se aqui para a exaustão dos *items* antes de proceder a uma escolha. Em certos casos, adopta-se uma posição menos rigorosa, mas no entanto muito estudada (ver exemplo do grupo de Popham, relativo à biologia).

Para cada micro-objectivo, precisam-se os comportamentos a adoptar e os critérios de êxito para cada indivíduo ou para cada grupo.

Exemplos:

- Para cada objectivo, o aluno deverá acertar pelo menos em sete *items* em dez.
- Para cada objectivo, 90% do grupo deve acertar pelo menos em sete *items* em dez.

Sublinhemos ainda que semelhantes controlos não concernem o método de ensino. Seja ele qual for, interessa verificar em certos momentos em que medida se atingem os objectivos fixados, ou, ainda, os conhecimentos prévios que os alunos dominam antes de encarar um novo objectivo.

Como os micro-objectivos operacionais constituem os modelos directos dos *items*, remetemos o leitor os exemplos para propostos nas págs. 128-130.

A teoria da construção dos testes centrados sobre os objectivos está sempre em pleno desenvolvimento¹.

Como o fim prosseguido é essencialmente diagnóstico, o importante, para a prática escolar, é organizar quadros reca-

¹ Ver particularmente Y. TOURNEUR, *Application d'indices et de tests statistiques à l'évaluation de maîtrise*, Louvain, Colloque de l'Association Internationale de Pédagogie expérimentale de Langue Française, 1973.

² J. PAQUAY-BECKERS, *D'une philosophie de la compensation à une pédagogie de maîtrise*,

pitulativos que mostrem claramente a situação de cada aluno. Saberemos assim onde intervir.

Apresentamos um exemplo correspondente ao objectivo «Estudo das qualidades estáticas de conjuntos de objectos», apresentado na pág. 132.

QUALIDADES ESTÁTICAS DOS CONJUNTOS DE OBJECTOS (UMA QUALIDADE)

4 de Outubro de 1973

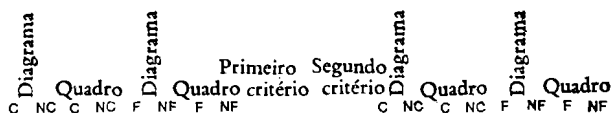
$M_{2b} E_{2a} O_1$

$- E_{2a} O_2$

$M_{2a} E_{2b} O_2$

Legenda

- × êxito total
- êxito parcial
- reprovação



Sália	X	X	X	X	X	X	+1	-2	X	X	X	X	X	X	X	X	X	X	X	X
Sabina	X	-4	X	-4	X	-	-2	-	X	-	·	-	X	X	X	X	X	X	X	X
Isabel	X	X	X	X	X	-	-	-	X	X	·	-	X	-	X	X	X	-	X	-
Joana	X	X	-1	X	X	-1	X	-1	X	X	+1	-	X	X	X	X	X	X	X	X
Fáí	X	X	X	X	-	-	-	-	X	-	-	-	X	-	X	X	X	-	X	-
Fátima	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Antónia	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Júlia	X	X	X	X	X	-	X	-3	·	-	·	-	X	-	X	X	X	X	X	X
Manuela	X	X	X	-	-	-	-	-	X	X	-	-	X	X	X	X	-	-	-	-

N. B. — Os números das quadrículas designam o número de objectos que a criança se esquece de classificar, quando levam o sinal menos (—), ou o número de objectos que a criança junta quando levam o sinal mais (+).

CAPÍTULO 4

CONTROLO DA FIDELIDADE DO EXAME

Muitas das medidas que temos vindo a considerar até agora tendem a assegurar uma melhor justiça escolar.

O controlo da fidelidade obedece à mesma preocupação. Teoricamente, um mesmo exame (escrito ou oral), feito muitas vezes sem que o aluno tenha tempo de aprender coisas novas, deveria conduzir ao mesmo resultado. Gostaríamos que fosse tão *fiel* como o metro que, bem utilizado, mede sempre o mesmo comprimento, com a possibilidade dum pequeníssimo erro.

Evidentemente que não permite um bom controlo da fidelidade mandar realizar a mesma prova diversas vezes. A primeira vez provoca uma reflexão geradora de aprendizagem. Em seguida, o aluno continua a pensar no problema, verifica a exactidão das respostas, adquire informações complementares, etc. Numa palavra, temos de recorrer a um outro meio ou, mais exactamente, a um conjunto de meios favoráveis à fidelidade.

Admitindo que se tenha conseguido a fidelidade da correcção, resta garantir a fidelidade das respostas.

1. Evitar toda a ambiguidade nas perguntas

Um pergunta que não é clara, a ponto de se prestar a muitas interpretações diferentes, priva o exame, de antemão, da sua fidelidade. Efectivamente, nada permite prever com alguma

certeza qual dos sentidos poderia ter sido dado por um mesmo indivíduo ou por um mesmo grupo, em diversas ocasiões.

Considerámos, na página 98, um certo número de meios para assegurar a clareza das perguntas.

2. Perguntas em número suficiente

Se pusermos apenas um pequeno número de perguntas, deixamos vastas zonas de matérias inexploradas. Ora pode acontecer que o aluno não tenha compreendido todas as matérias ou tenha negligenciado partes do curso, especulando com a probabilidade ou com a preferência manifesta do professor. Segundo a parte em que incidirem as perguntas, irá obter um resultado brilhante ou um fracasso.

O único meio de evitar semelhante aventura é abranger toda a matéria. Por outros termos, o exame deve conter uma *boa amostra* do conjunto.

Em princípio, quanto maior for o número de perguntas mais a fidelidade aumenta.

3. Um controlo matemático

Para exames importantes, será desejável obter ainda maior garantia.

a) *Método pares-ímpares*

Se o número de perguntas for bastante elevado, o seguinte procedimento simula uma situação em que o mesmo aluno faria duas vezes o mesmo exame, praticamente ao mesmo tempo.

a) Numeradas as perguntas, dividem-se em dois grupos: pares e ímpares.

- b) Organizam-se dois quadros paralelos dos resultados e calcula-se a sua correlação (r); noutros termos, avalia-se numericamente a relação que existe entre os dois grupos¹.
- c) Os dois exames artificialmente criados correspondem cada um à metade do original. Em virtude do que vimos no parágrafo 2, são portanto menos fiéis e a correlação não tem em conta a situação real.

Uma fórmula simples permite operar a correcção necessária. O índice de fidelidade obtém-se finalmente pela fórmula

$$\frac{2r}{1+r}$$

Quanto mais o resultado se aproximar de 1, tanto mais elevada será a fidelidade.

Em geral exige-se pelo menos 0,80.

Quando se prepara o exame mediante um teste prévio, a insuficiente fidelidade é combatida com o aumento do número de perguntas.

Se o controlo da fidelidade se faz após o exame efectivo, uma fidelidade demasiado fraca incitará os professores a um suplemento de prudência na interpretação dos resultados e a um maior cuidado na preparação de exames futuros.

b) *Duas formas paralelas*

O controlo matemático é o mesmo que no método *pares-ímpares*, mas neste, os examinadores preparam duas séries completas de perguntas, unanimemente julgadas equivalentes, a partir do mesmo conjunto de objectivos.

Se a população é bastante numerosa, metade, escolhida ao acaso (por exemplo, segundo a ordem alfabética dos nomes) recebe a primeira parte e a outra metade, a segunda. Para

¹ O método do cálculo da correlação, é explicado em todos os manuais elementares de estatística.

populações pouco numerosas, convidam-se os alunos a responder às duas séries de perguntas alguns dias depois. Esta segunda solução é evidentemente muito menos rigorosa.

4. Repetição da classificação

Não basta a fidelidade das respostas. Ela deve ser acompanhada pela fidelidade da classificação. Esta controla-se quer convidando um mesmo professor a classificar duas vezes os mesmos trabalhos com intervalo de alguns dias ou de algumas semanas, quer comparando as notas dadas por vários examinadores a uma mesma prova.

CAPÍTULO 5

CONTROLO DA VALIDADE

A fidelidade dum exame não garante em nada a sua validade. Um metro mal construído, que medisse cinco centímetros a mais, indicaria o mesmo comprimento em cada medida executada, mas conduziria a uma conclusão falsa.

Numa palavra, tornar válido um exame é provar que ele mede efectivamente aquilo que se propôs medir. É fundamental obter esta certeza. Com efeito, se a conclusão for negativa, são vãos todos os esforços desenvolvidos para a organização e efectuação dos exames ou, pelo menos, desviados do seu objecto.

Ora, a validação não é coisa fácil. Ao fazer o cômputo sobre os nossos conhecimentos a respeito da validade, R. Cox concluiu, em 1969, que a investigação mal afluou as questões fundamentais¹. O problema merece a maior atenção e prudência dos professores.

Segundo a natureza retrospectiva ou prospectiva do exame, distingue-se a validade do conteúdo e a validade de previsão.

1. Validade do conteúdo

Interessa sobretudo para os exames destinados a fazer o balanço das aquisições e por conseguinte também dum ensino.

¹ R. COX, *Reliability and Validity of Examinations*, in J. LAUWERYS e D. SCALON, *Examinations*, o. c., p. 43.

Essa finalidade tinha sido indicada e os objectivos escolhidos. Ter-se-ão atingido?

Nos exames que incidem sobre longas etapas de aprendizagem, é evidente que não se pode fazer um interrogatório pormenorizado sobre toda a matéria. Não só devemos localizar os pontos principais como ainda ter em conta a sua importância relativa. Se forem satisfeitos esses dois critérios, as perguntas poderão suscitar respostas *representativas* do conjunto das aquisições, da competência total.

Por outras palavras, a validade do conteúdo depende da qualidade e da adequação da amostragem dos conhecimentos e das capacidades.

Já vimos que, num ensino bem concebido, os objectivos a atingir, e portanto a controlar na altura do exame, foram definidos no início do ano.

Mas então tratava-se de intenções que as circunstâncias podem ter modificado. Antes de tentar fazer o balanço do trabalho dos alunos, o professor fará pois em primeiro lugar o seu. Se faltar este alinhamento, os exames terão apenas uma validade illusória ou, pior ainda, uma validade de fachada, destinada a desobrigar o professor aos olhos dos seus superiores. Assim se explica que perguntas incidam sobre matérias apenas afluadas, embora figurando no programa oficial, ou sobre capacidades intelectuais de nível elevado que não foram desenvolvidas paciente e sistematicamente durante o ano. Quantos malogros escolares não se devem às lindas perguntas «de inteligência» feitas no final dum ensino que não a cultivou?!

Se a definição precisa dos objectivos permite tornar uma validação de conteúdo muito melhor que em tempos idos, não deixa no entanto de ser muito limitada no estado actual dos nossos conhecimentos. Em primeiro lugar, porque é muito difícil obter a certeza de que determinado comportamento real é bem representativo de determinada característica que se pretendia cultivar. Depois, porque, mesmo que o ponto de vista seja correcto no início, as vicissitudes da classificação podem trair sempre as intenções.

Um professor de história ou de geografia, escreve P. Vernon¹, que mediante um exame escrito tenta avaliar a compreensão profunda dessas disciplinas, atribui muitas vezes uma boa parte dos pontos a uma reprodução exacta dos factos nos seus pormenores, à repetição das suas teorias favoritas, à extensão das respostas, à intelegibilidade e elegância do estilo com que foram expressas. Num exame oral, quantas coisas dependem da inteligência social, do controlo de si mesmo e, particularmente, da aplicação do aluno para criar, pela palavra e pela atitude, um bom contacto com o examinador!

Que fazer na prática?

Primeiramente exigir que o professor, para cada pergunta de exame que redige, indique o objectivo numa ficha especial

Em seguida submeter as perguntas e o esquema de classificação a vários educadores experimentados (dos quais um pelo menos pertencerá a um grupo de disciplinas diferente do considerado) que julgarão independentemente da validade do conteúdo das perguntas. Em caso de desacordo grave, a pergunta deve ser rejeitada².

A experiência indica que actualmente devemos encarar só um pequeno número de objectivos gerais concebidos de maneira bastante ampla. Caso contrário, os professores não conseguirão mais pôr-se de acordo. Já dissemos atrás que os nossos conhecimentos em matéria de validade são ainda muito pobres.

Pela mesma razão, nenhum acordo poderá obter-se em certos casos a propósito de questões fundamentais. Claro que não é motivo suficiente para pôr fim ao ensino da matéria ou da disciplina incriminada, mas deve ser claramente indicada a impossibilidade de acordo.

¹ P. VERNON, *Types of Examinations*, in LAWREY e D. SCANLON, *Examinations*, p. 43.

² Pode calcular-se um coeficiente de concordância entre as opiniões pela fórmula de Kendall e escolher-se um limiar de aceitação de comum acordo.

Podemos encontrar essa dificuldade, por exemplo, a propósito do controlo sistemático da realização dos objectivos atribuídos às línguas antigas, a certas matérias da matemática, etc.

Verificamos de novo que, procurando aumentar o rigor científico dos exames, vamos pôr em causa tipos de ensino. Mas só assim poderá desaparecer o feiticismo pedagógico.

II. Validade de previsão

Parece de longe a mais simples de controlar, porque não exige compreensão profunda dos fenómenos. Mesmo que não se explique a relação existente entre o êxito em determinado teste e o resultado brilhante em certo domínio, o facto pode ser facilmente observado.

Quer se trate dum exame que dá acesso a um novo ciclo de estudos ou que autorize o exercício de determinada profissão, basta seguir os indivíduos durante um certo tempo para saber se o prognóstico formulado a partir das notas atribuídas se verifica ou não.

Observa-se com surpresa que uma tal verificação raramente se faz. J. C. Passeron vê nisso o sinal de que o exame se destina menos a medir a capacidade do que a servir as necessidades *sociais* das classes privilegiadas¹.

Esta afirmação merece certas reservas mas, como vimos na primeira parte, não é totalmente desprovida de fundamento.

Para não ser um logro, a verificação sistemática da validade de previsão exige o controlo rigoroso das variáveis, sem o que as variáveis ocultas falsearão as conclusões. Por exemplo, ainda recentemente se argumentava a favor do latim, fazendo a estatística do número de estudantes excepcionais que tinham feito estudos secundários clássicos. Ora, esta observação não prova

¹ J. C. PASSERON, *Sociologia des examens*, oc. c., p. 7.

nada, a não ser que, tradicionalmente nos nossos países, os alunos mais inteligentes são orientados a partir dos doze anos para esse tipo de estudos.

A modo de síntese, vamos buscar a recapitulação seguinte ao *Examination Bulletin* n.º 3¹.

FACTORES DE VALIDADE DUM EXAME

1. Identificação adequada dos objectivos cuja realização deve ser verificada pelo exame escrito, avaliação do trabalho do ano, prova prática, etc.
2. De entre toda a gama de objectivos assim identificados, selecção daqueles em que vai concentrar-se a verificação.
3. Avaliação eficaz da adequação do conteúdo e da estrutura do exame ao fim em vista.
4. Relação clara entre cada pergunta e os objectivos do ensino.
5. Elaboração dum esquema de classificação e redacção das orientações para os examinadores em função dos objectivos. Deve também assegurar-se um nível suficiente de fidelidade.
6. Bom conhecimento das capacidades dos candidatos.
7. Disposições dos examinadores a terem em conta juízos independentes do seu e os dados objectivos que lhes sejam fornecidos.
8. Comparação com provas anteriores cuja validade já se provou.

¹ Londres, H. M. S. O., 1964, pp. 19-20.

CAPÍTULO 1

POSIÇÃO DO PROBLEMA

1. Definição

A *moderação*, em docimologia, teve primeiramente como objectivo atenuar os excessos de severidade ou de generosidade de alguns examinadores.

Em sentido lato, designa hoje o conjunto de medidas tomadas para tornar comparáveis as notas de exames internos, e, portanto, para unificar o seu significado ao nível das diferentes classes do mesmo tipo numa escola, num grupo de escolas, em estabelecimentos similares duma região ou dum país.

As investigações mais cuidadosas sobre a moderação são devidas sem dúvida aos docimologistas britânicos. Num sistema completamente descentralizado como o da Grã-Bretanha, os programas e os métodos das escolas primárias ou secundárias podem diferir consideravelmente. É assim, durante muito tempo, exames externos, isto é organizados fora das escolas por comissões especiais, tiveram de sancionar os estudos nos momentos cruciais da escolaridade. É conhecido o famoso 11 + *Examination*, exame de entrada no ensino secundário, que decide, de modo particular, da admissão ao ensino geral clássico e o C. S. E. (*Certificate of Secondary Education*), antes da entrada no ensino superior. Este último exame, por exemplo, fornece a cada aluno um documento que indica o nível em cada uma das disciplinas em relação a normas nacionais.

Para admitir um aluno a uma secção determinada, as universidades formulam as suas exigências particulares. Num

caso, são exigidas notas muito boas (A) em Matemática, Biologia e Língua Materna; noutro caso, aceitam-se também notas boas (B), etc.

Este sistema de exames externos é, às vezes, muito severamente criticado, porque limita de maneira grave a liberdade pedagógica dos professores e das autoridades locais.

É por isso que já há alguns anos se tenta substituí-los por exames internos cujos resultados se tornam comparáveis mediante processos de moderação local, regional, depois nacional, já ajustados experimentalmente.

Duma maneira sumária, a razão é dupla. Primeiramente pelo que diz respeito aos indivíduos, alunos e pais, que têm o direito de saber o nível «real» dos resultados escolares, antes de decidir da orientação posterior no domínio dos estudos ou da profissão. Quem é que se julgaria qualificado para os jogos olímpicos pelo facto de ter ganho uma corrida organizada entre alguns amigos? Um juízo desfavorável não é, aliás, tanto uma condenação sem apelo como um diagnóstico de fraqueza ou de doença. Teremos de saber a tempo que há um problema para lhe buscar as causas, depois os remédios, se existirem.

Em segundo lugar no que respeita à comunidade. Assim como um consumidor não deve ser enganado a respeito da mercadoria, assim também a sociedade não pode ser obrigada a pagar muito caro os malogros dum estudante entrado na universidade com base num certificado inválido ou a confiar uma função a alguém que não a merece. No dia em que os exames do ensino normal forem rigorosamente controlados, poderemos ter a esperança de não mais ver confiar gerações de crianças a certos professores cujo conhecimento da língua materna, designadamente da ortografia ¹, é mais que insuficiente.

¹ Não temos a superstição da ortografia, mas admitimos como evidente que aquele que a vai ensinar deva ao menos conhecê-la.

2. Moderar não é arregimentar

Poder-se-ia rezear que o desejo de tornar os resultados escolares comparáveis viesse a dar como resultado um ressurgir dos antigos constrangimentos. Não esquecemos esse inspector de tradição napoleónica que, ao consultar o relógio, julgava poder afirmar: «Neste momento, ensina-se determinada lição em todas os quintos anos primários de França». Todos sabemos as deficiências dum ensino arregimentado.

Mas a necessidade duma plasticidade pedagógica não impede que, para além dos arranjos de circunstância, cada ciclo de estudos prossiga objectivos fundamentais, comuns a todos: aquisição de técnicas de base ou de conhecimentos e capacidades admitidas como essenciais. Professores e moderadores devem interrogar-se e entender-se a propósito dessas aprendizagens essenciais.

3. Moderação voluntária ou imposta?

Normalizar todos os exames em todas as etapas dos estudos não parece nem desejável nem possível. Pouco desejável, porque os alunos progridem a ritmos às vezes muito diferentes e dificilmente se acomodam a uma divisão rigorosa em períodos e anos escolares. Impossível, porque a moderação, mesmo com os métodos económicos que vão ser apresentados, continua um trabalho árduo, tanto na preparação como na execução.

A moderação deve ser imposta nos momentos decisivos, acima de tudo na altura da atribuição de diplomas ou de certificados de fim de ciclo.

Nos casos restantes, deveria ser deixada a decisão ou aos professores individualmente ou aos directores dos estabelecimentos de ensino ou às autoridades locais.

4. A moderação começa no início do ano escolar

Como dissemos, é preciso que professores e moderadores tenham conseguido um acordo sobre os objectivos do ensino e sobre alguns grandes princípios de classificação para que os exames sejam comparáveis.

Afirmações vagas tais como «O aluno deve saber exprimir-se correctamente por escrito» não têm utilidade. A capacidade desejada deve traduzir-se em termos de comportamentos concretos que terão todo o interesse em ser encorporados nas escalas de avaliação descritivas. Se se tratar da expressão escrita várias perguntas se põem.

Como graduar as exigências em função da idade dos alunos e do tipo de escola?

A originalidade será recompensada? Como? Como poderá ser identificada?

Que importância atribuir à ortografia?

Será considerada a extensão dos trabalhos de composição?

Exigir-se-á uma extensão mínima, abaixo da qual haverá pontos descontados? Quantos?

Que importância atribuir à riqueza do vocabulário?

Como defini-la? Quais serão as exigências sintácticas?

Os matemáticos podem afinar as suas violas. Para nos limitarmos a um único exemplo, se certos professores são muito severos quanto aos erros nas operações, até mesmo para os erros de ortografia na solução dos problemas, outros colegas consideram esses aspectos secundários donde resulta que notas iguais podem abranger realidades muito diferentes.

Os programas oficiais podem desempenhar também um papel importante, precisando o mais possível as finalidades a atingir em cada um dos escalões. A eficácia desses dados de base é tanto maior quanto mais centralizado for o sistema escolar.

Inútil criar ilusões: os objectivos comportamentais das diferentes disciplinas não se definem nalgumas horas de meditação. Antes de chegarem a um resultado satisfatório, deverão trabalhar longamente, comissões em que professores, inspectores e investigadores unirão os seus esforços. Muitas perguntas poderão, aliás, ficar provisoriamente sem resposta.

A definição dos objectivos será objecto duma investigação permanente, não apenas porque poderão adquirir precisão com o avanço da psicologia da aprendizagem, mas ainda porque os próprios objectivos irão evoluindo.

As instruções relativas aos exames ir-se-ão precisando paralelamente. Ao lado das séries de objectivos gerais e especiais, os programas escolares do futuro determinarão claramente a quantidade e o tipo de perguntas a fazer nos exames e fornecerão indicações sobre as linhas de força dos esquemas de classificação.

5. Não há possibilidade de comparação sem fidelidade elevada

Não é concebível a comparação de dois resultados de exames que, tomados isoladamente, seriam eminentemente flutuantes.

O problema da fidelidade foi discutido na parte consagrada à construção do exame. Não vamos insistir nele.

6. Poderemos fiar-nos nos testes?

Nos sistemas de moderação que vamos examinar, ocupam um lugar importante os testes de inteligência ou de conhecimentos. Em certos momentos, fornecem os pontos de referência aos quais irão ajustar-se os outros resultados; noutros casos, dá-se-lhes tanto valor só a eles como ao trabalho do ano e ao exame final.

Agora que já vimos com que rigor e precauções os bons testes de conhecimentos são construídos, não poderá surpreen-

der-nos a confiança atribuída a esses instrumentos. Além disso, vimos, a partir da primeira parte deste livro, que se os professores julgarem bem os alunos, um simples teste verbal, administrado em menos duma hora, permitirá também uma previsão quase tão segura:

	Correlação com resultados do aluno	
	2 anos depois	3 anos depois
Prognóstico do professor	. 821	. 748
Teste verbal	. 796	. 722

Vejamos ainda alguns resultados de investigações a confirmar essas observações.

A médio prazo, um teste de inteligência ou, melhor ainda, uma combinação de *scores* de diversos testes de inteligência, são melhores indicadores do êxito escolar do que os resultados de exames. Essa superioridade foi demonstrada repetidas vezes, sobretudo por investigadores anglo-saxões:

- 1) Emmet¹ mostra que um teste de inteligência verbal permite melhor prever os resultados, após dois ou três meses de ensino secundário geral, do que os exames de língua materna e de aritmética avaliados pelos professores.
- 2) Emmel e Wilmut² fizeram depois uma demonstração do mesmo modo convincente do valor de previsão do teste de inteligência, cinco anos após.
- 3) Wrigley³ confirmou esses resultados e mostrou que a previsão pode ser ainda melhor se for possível combinar

¹ W. EMMET, *An Inquiry into the Prediction of Secondary School Success*, London Univ. of London Press, 1942.

² W. EMMET and F. WILMUT, The Prediction of School Certificates Performance in Specific Subjects, in *British Journal of Educ. Psychol.*, 22, 1952, 52-62.

³ J. WRIGLEY, *The Relative Efficiency of Intelligence and Attainment Tests as Predictors of Success in Grammar Schools*, in *British Journal of Educ. Psychol.*, 25, 1955, 107-116.

os resultados dum teste de inteligência e de testes de conhecimentos estandardizados.

- 4) Na Bélgica, todos os orientadores conhecem o valor de previsão dos *scores* verbais e de raciocínio do Teste P. M. A. de Thurstone, o famoso $2V + R$, para o ensino secundário geral.

Nós não dispomos de investigações similares para o ensino técnico ou para estruturas de ensino talvez melhor adaptadas à civilização do ano 2000. É muito possível que devessem ser utilizados *scores* doutros tipos de testes.

CAPÍTULO 2

ALGUNS SISTEMAS DE MODERAÇÃO DOS EXAMES

I. Por referência a um teste

A. A fórmula mais liberal: O sistema sueco de moderação por disciplina a partir de testes de conhecimentos

Desde o fim da segunda guerra mundial, a Suécia adoptou um sistema de moderação, simples e facultativo, aplicável por cada professor na sua classe, ao longo da escolaridade.

Todas as notas dos professores são atribuídas segundo uma escala de 7 graus ¹.

Sugere-se que a percentagem de alunos com uma nota determinada seja sumariamente a seguinte ², ficando o professor com a liberdade de ter em conta as características da sua classe (por exemplo, aluno excepcional, ou quase todos os alunos médios, etc.):

Notas	1	2	3	4	5	6	7
Percentagem	1	6	24	38	24	1	1

¹ Referimo-nos à exposição sobremaneira clara de S. HENRYSSON, *The Swedish System of Equalising Marks*, in *Educational Research*, VI, 2. Fev. de 1964, 156-160.

² Distribuição normal (ver curva de Gauss).

Além disso, todos os anos um organismo central, em que colaboram especialistas e professores experimentados, elabora uma bateria de testes de conhecimentos incidindo sobre as disciplinas principais do programa. Esta bateria é aferida na base duma amostra nacional representativa. Os resultados serão distribuídos em 7 classes, na proporção aconselhada pelas notas dos professores.

Se o professor administrar o teste nacional na sua classe, disporá de notas directamente comparáveis às notas que atribuiu. É fácil pois ajustar estas últimas.

Demos um exemplo concreto proposto por S. Henrysson:

<i>24 alunos</i>			
<i>Notas</i>	<i>Distribuição das notas preliminares atribuídas pelo professor</i>	<i>Distribuição das notas obtidas no teste</i>	<i>Distribuição das notas depois de ajustadas</i>
7	—	1	—
6	—	4	4
5	4	4	4
4	4	9	10
3	14	5	4
2	2	1	2
1	—	—	—
<i>Média</i>	3,42	4,33	4,17

Em primeiro lugar, o professor inscreve na 2.^a coluna as notas que atribuiu. A média (3,42) dá uma primeira indicação, se a média nacional for conhecida. (Na Suécia, é de 4 para a 6.^a primária. Parece portanto que ali o professor é demasiado severo).

Vejamos agora os resultados obtidos no teste pelos mesmos alunos:

<i>Score bruto</i>	<i>Nota correspondente</i>	<i>Número de alunos que obtêm essa nota</i>
94-100	7	1
84-93	6	4
67-83	5	4
46-66	4	9
30-45	3	5
22-29	2	1
0-21	1	—

Estes resultados são dados no primeiro quadro, coluna 3.

A média é 4,33, o que confirma a primeira impressão de severidade.

Sem mudar a ordem da classificação inicial, o professor rectifica as suas notas (coluna 4).

O sistema sueco é sedutor por várias razões:

- 1.º Os professores conservam a sua inteira liberdade:
 - a) de recorrer ou não ao teste (quase todos o empregam);
 - b) de ter em conta os resultados;
 - c) de comunicar os resultados dos testes aos alunos, aos outros professores, ao director, aos pais.

Deste modo, o teste torna-se simplesmente um instrumento de ordem prática. Reduz-se assim ao mínimo a tentação de preparação intensiva para realizar uma prova.

- 2.º Os construtores de testes procuram abranger uma ampla gama de objectivos constituindo uma amostragem válida de todo o programa.

É evidente que os professores prestam uma atenção particular às matérias tratadas nos testes. Dispomos assim dum meio eficaz de sensibilizar o docente para certas inovações.

Note-se que um sistema semelhante foi experimentado em Genebra, a partir de 1965, na 5.^a e 6.^a classe da primária (10-12 anos), para a ortografia e a aritmética ¹.

De entre as críticas formuladas contra este sistema de ajustamento ², referiremos as seguintes:

1. Se um professor tiver julgado um dos alunos da sua classe com severidade excessiva, a referência à prova nacional não reparará a injustiça.
2. A prova de referência pode não medir as mesmas dimensões que as notas.
3. A elaboração da prova de referência pode ainda ser defeituosa: fidelidade demasiado baixa, ...

B. Sistema de moderação imposto por disciplina a partir dum teste de conhecimentos

A partir dos mesmos dados de base da Suécia (exames internos e testes de conhecimentos para as disciplinas principais, aferidas nacional ou regionalmente), o seguinte sistema directivo pode dar bons resultados.

Uma comissão de moderação nacional ou regional toma a seu cargo um certo número de escolas.

Essas escolas enviam-lhe os resultados das provas preparadas livremente pelos professores e os *scores* obtidos nos testes de conhecimentos.

Em caso de diferença para mais ou para menos, igual ou superior a dois desvios-padrão (por exemplo) entre as médias

¹ Ver S. ROLLER, Le problème de l'attribution des notes scolaires. Essai de solution. Genève, in *Docimologie et éducation*, número especial da revista *Les sciences de l'éducation*, avril-septembre 1969, pp. 66 seg.

² W. ANGOFF, Can Usual General Purpose Equivalency Tables be Prepared for Different College Admission Tests, in A. ANASTASI, Ed., *Testing Problems in Perspective*, Washington, ACE, 1966, pp. 251-264.

T. BACHER, *La normalisation des notes*, o. c., p. 63.

dos exames e as médias dos testes, a escola receberá a visita dos moderadores.

A sua missão não é dizer ao director como deve conduzir a escola, nem ao professor como fazer o curso, mas sim chamar a atenção para um facto e procurar descobrir, em perfeita colaboração com a escola, a explicação e, se possível, o remédio.

Se o desacordo subsistir, o direito de ajustar as notas pode ser dado às autoridades regionais, mantendo a escola o direito de interpor recurso.

C. Um sistema de selecção a partir dum teste de inteligência

Já observámos que em geral os professores avaliam bem os seus alunos, mas que tendem a relativizar os seus juízos tendo em conta o nível global da classe.

O seguinte sistema, elaborado pela Fundação Nacional Inglesa para a Investigação Educacional¹ com vista à admissão ao ensino secundário geral clássico, permite *seleccionar*, tendo em conta juízos feitos independentemente pelos professores na sua aula.

1. O professor classifica os alunos por ordem de mérito, isto é, segundo a sua avaliação global da probabilidade de êxito no ensino secundário geral. Muitos alunos poderão ser classificados *ex aequo* (classificação 1),
2. Os alunos fazem um teste de inteligência verbal (2).
3. Os *scores* do teste são classificados por ordem decrescente (classificação 3).

¹ Ver D. PIDGEON e A. YATES, o. c.,

4. As classificações (1) e (3) são postas lado a lado.
5. O *score* de inteligência que acompanha o nome do aluno é considerado como *score* de avaliação pelo professor (juízo ajustado).

Classificação do professor (1)	Score obtido pelo aluno no teste verbal (2)	Classificação dos scores por ordem decrescente (3)	Juízo ajustado (4)
A	121	132	132
B	120	128	128
C	132	121	121
D	128	120	120
E	100	106	106
F	106	100	100
G	94	100	97
H	82	96	97
I	96	94	97
J	100	86	86
K	78	82	78
L	79	79	78
M	86	78	78
N	73	73	78
O	65	65	65
		Total 290	
		Total 312	

Pode ver-se que a classificação a que por fim se chega em nada modifica a ordem inicialmente escolhida pelo professor. Mas desta vez dispomos dum meio de comparação entre escolas.

Utilização dos critérios ajustados

Exemplo: Num liceu, pretende-se recrutar, para o 1.º ano da secção clássica, 25 alunos provenientes de três escolas primárias diferentes.

Critérios ajustados

	Escola primária I	Escola primária II	Escola primária III
130	132		130 x 2
125	128	125 124	126 125
120	121 120	121 119 117	123 120
115	115 x 3	115	116
110	112	110	114 113 110
105	106	107 104	
100	100		

Para seleccionar 25 alunos, corta-se neste caso em 110. Se, no corte de 110, dos *ex aequo* tivessem elevado para mais de 25 alunos a categoria dos seleccionados, a mais elementar justiça exigiria que se admitissem todos.

Traçar uma linha de demarcação como acabámos de fazer é uma fonte de injustiça por uma outra razão. Não será apenas o acaso o responsável pela colocação dum certo número de examinandos acima ou abaixo do limite?

O psicometrista está armado para vencer essa dificuldade: fixa uma margem de segurança de três vezes o erro médio que uma simples fórmula permite calcular a partir do coeficiente de fidelidade do teste utilizado ¹.

¹ Erro *standard* = $\sigma \sqrt{1-r}$

em que σ = desvio-padrão dos scores

r = coeficiente de fidelidade. Este coeficiente é indicado geralmente no manual que acompanha o teste.

II. Moderação por recurso a um banco de itens

No caso de os professores utilizarem perguntas sob a forma de *itens* de escolha múltipla, parece extremamente promissor a técnica seguinte actualmente experimentada na Grã-Bretanha¹. Todavia, exige a existência dum serviço de investigação, necessidade de que nenhum educador consciente poderá hoje duvidar.

1. Os professores indicam numa folha quadriculada os objectivos considerados.
2. Envia esta folha ao serviço de investigação, e ao mesmo tempo, os *itens* que redigiram e, eventualmente, já prestatos por eles.
3. O serviço examina esses *itens* e, de acordo com as suas possibilidades, põe à prova alguns deles; calcula-se a sua dificuldade e poder discriminativo.
4. O serviço devolve ao professor:
 - a) os *itens* examinados
 - b) *itens* complementares cujos índices de facilidade e de eficácia são conhecidos para uma determinada população: cidade, região, país, ...

Esses *itens* fornecerão um ponto de comparação a partir do qual poderá ajustar-se o resultado obtido pelos outros.

É interessante notar que os ingleses, depois de se terem mantido estreitamente acantonados na redacção de *itens* de resposta fechada, por escolha múltipla, começam a propor perguntas semi-abertas. Efectivamente, trata-se de perguntas abertas, formuladas com todo o cuidado, para as quais se identificaram experimentalmente os grandes tipos de respostas, em função das quais se propõem tabelas de correcção.

A organização dum banco de *itens* é trabalhosa durante os primeiros anos. Com o tempo, a provisão de *itens* aferidos

¹ Ver D. PIDGON e A. YATES, *o. c.*

torna-se tal que o trabalho se aligeira e adquire uma grande flexibilidade e uma grande rapidez de funcionamento.

A dificuldade das perguntas pode variar a prazo relativamente reduzido, segundo a flutuação dos programas e do ensino. O índice de dificuldade deverá pois ser ajustado continuamente em função dos resultados observados no momento da utilização de cada *item*.

III. Processo de equilibração

Um sistema completo de moderação em Inglaterra

O sistema que vamos apresentar agora é também devido a estudiosos ingleses. Quanto é do nosso conhecimento, não há outro mais completo. Diz respeito à preparação do exame em colaboração com as escolas, à classificação, ao ajustamento das notas e dos resultados finais.

O fim em vista é *aperfeiçoar os exames internos*, ao ponto de lhes poder atribuir o papel desempenhado até ao presente pelas grandes provas externas.

Para chegarmos a uma moderação nacional, procede-se por etapas. Ao nível local, agrupa-se um certo número de escolas que prosseguem uma mesma finalidade. Ao nível regional depois, segundo uma técnica que não difere fundamentalmente da técnica do nível local. Para passar ao nível nacional, acrescenta-se à técnica de moderação propriamente dita uma amostragem final cujo estudo técnico não viria aqui a propósito.

Para o nosso país (Bélgica), sugerimos que algumas escolas com o mesmo espírito comecem por fazer uma experiência voluntária do sistema. Por essa razão, vamos concentrar o nosso estudo sobre o processo de base.

A moderação inter-escolas só se faria nos momentos cruciais da escolaridade e para algumas disciplinas principais.

Todavia, nada impede os professores duma mesma escola, regendo os mesmos cursos, de utilizarem espontaneamente o mesmo método de correcção.

A. Preliminares

Já fizemos alusão à concordância a estabelecer quanto aos objectivos e à construção do exame. Não vamos insistir.

Cada escola escolherá um professor que desempenhará dois papéis:

- 1.º coordenador dos exames na sua escola;
- 2.º membro da comissão de moderação inter-escolas.

Um retrato ideal do moderador não existe. Parecem desejáveis as seguintes qualidades¹:

1. Maturidade geral e boa experiência pedagógica;
2. contacto frequente com alunos do nível e do tipo de escola a que dizem respeito os exames;
3. capacidade de formular claramente os seus critérios e os seus juízos;
4. bondade, mas firmeza;
5. capacidade de discutir sem paixão, sem criar tensão;
6. aptidão para compreender algumas técnicas de análise estatística.

Em cada escola, as perguntas dos exames serão redigidas com grande liberdade de movimento. Só a folha quadriculada com os objectivos principais, o número e o volume das perguntas devem ser respeitados segundo o acordo estabelecido.

Ao princípio pelo menos, é bom submeter as perguntas à comissão de moderação que verifica a unidade de conjunto.

¹ MATHER, o. c., p. 67.

B. Os professores classificam os seus exames

C. Nova correcção duma amostra

Três princípios dominam o trabalho:

- 1.º A intervenção das escolas nos exames deve ser tão discreta quanto possível. Só os casos de divergências importantes merecem atenção.
- 2.º Os professores que corrigem devem ter aproximadamente a mesma severidade, apreciar as mesmas qualidades e estar de acordo sobre o significado da classificação final.
- 3.º As amostras para correcção devem ser pequenas e o trabalho estatístico tão simples quanto possível.

Em cada escola e para um mesmo tipo de exame, tomam-se ao acaso 20 exemplares corrigidos.

Seja, por exemplo, o caso de 12 liceus¹. Cada um envia um delegado moderador à comissão que recebe assim 12 maços de 20 exercícios.

A primeira operação tem como fim determinar a equivalência dos moderadores sob três pontos de vista:

- 1.º *Severidade* — cujo grau é posto em evidência pela média.
- 2.º *Discriminação* — classificação exageradamente desvoluta ou exageradamente prudente. É revelada pela dispersão ou margem de variação das notas.
- 3.º *Conformidade* — um mesmo aluno é classificado de igual modo por todos os professores? É revelada pela correlação entre duas séries de notas.

Para verificar a concordância entre moderadores, sob estes três pontos de vista, os doze corrigem primeiro um mesmo maço de 20 exemplares.

¹ Exemplo tirado de *Examinations Bulletin n.º 5*, Londres, HMSO, 1965.

Os 20 exemplares são repartidos, sempre ao acaso, em cinco grupos de quatro. Em cada grupo, os alunos são classificados por ordem alfabética.

Em seguida, fazem-se as simples operações seguintes, pela ordem em que figuram no quadro:

ESTUDO DA CONCORDÂNCIA ENTRE OS MODERADORES

Notas dadas por 12 moderadores

Nome do candidato	MODERADORES												Média
	a	b	c	d	e	f	g	h	i	j	k	l	
A	4	4	4	3	4	4	3	1	4	4	4	4	4
B	2	2	2	2	3	2	1	1	2	2	2	2	2
C	6	5	5	5	5	5	5	5	5	5	5	5	5
D	2	1	1	2	2	3	3	2	2	3	2	3	2
Total	14	12	12	12	14	14	12	9	13	14	13	14	13
Margem de var.	4	4	4	3	3	3	4	4	3	3	3	3	3
E	3	3	3	4	3	5	3	3	3	3	3	3	3
F	1	1	1	1	1	1	1	1	1	1	1	1	1
G	5	6	6	6	5	3	6	6	6	6	6	6	6
H	3	4	4	4	3	4	3	4	4	2	4	4	4
Total	12	14	14	15	12	15	13	14	14	12	14	14	14
Margem de var.	4	5	5	5	4	4	5	5	5	5	5	5	5
I	2	2	2	2	4	3	2	3	3	3	3	3	3
J	2	1	2	1	3	2	2	2	2	2	2	2	2
K	5	5	5	6	5	5	5	6	5	5	5	4	5
L	1	2	2	1	2	2	2	3	2	2	3	2	2
Total	10	10	11	10	14	12	11	14	12	12	13	11	12
Margem de var.	4	4	3	5	3	3	3	4	3	3	3	2	3
M	4	2	3	4	3	4	5	5	5	4	5	4	4
N	4	3	4	4	4	5	4	4	4	4	4	4	4
O	5	4	4	3	4	3	4	4	4	5	4	5	4
P	5	6	5	5	5	5	5	5	5	5	5	5	5
Total	18	15	16	16	16	17	18	18	18	18	18	18	17
Margem de var.	1	4	2	2	2	2	1	1	1	1	1	1	1
Q	1	2	1	1	1	1	1	1	1	1	2	1	1
R	1	1	2	1	1	2	1	1	1	1	1	1	1
S	3	3	2	3	3	2	2	3	3	3	3	3	3
T	2	1	2	1	2	2	1	2	2	2	2	2	2
Total	7	7	7	6	7	7	5	7	7	7	8	7	7
Margem de var.	2	2	1	2	2	1	1	2	2	2	2	2	2
Total geral	61²	58	60	59	63	65	59	62	64	63	66	64	63¹
Soma das margens	15⁴	19	15	17	14	13	14	16	14	14	14	13	14³

Diferenças entre as notas dadas por cada moderador e média das notas dadas por todos os moderadores, a um mesmo candidato.

Nome do candidato	MODERADORES											
	a	b	c	d	e	f	g	h	i	j	k	l
A	0	0	0	-1	0	0	-1	-3	0	0	0	0
B	0	0	0	0	1	0	-1	-1	0	0	0	0
C	1	0	0	0	0	0	0	0	0	0	0	0
D	0	-1	-1	0	0	1	1	0	0	1	0	1
Margem de var.	1	1	1	1	1	1	2	3	0	1	0	1
E	0	0	0	1	0	2	0	0	0	0	0	0
F	0	0	0	0	0	0	0	0	0	0	0	0
G	-1	0	0	0	-1	-1	0	0	0	0	0	0
H	-1	0	0	0	-1	0	-1	0	0	-2	0	0
Margem de var.	1	0	0	1	1	3	1	0	0	2	0	0
I	-1	-1	-1	-1	1	0	-1	0	0	0	0	0
J	0	-1	0	-1	1	0	0	0	0	0	0	0
K	0	0	0	1	0	0	0	1	0	0	0	-1
L	-1	0	0	-1	0	0	0	1	0	0	1	0
Margem de var.	1	1	1	2	1	0	1	1	0	0	1	1
M	0	-2	-1	0	-1	0	1	1	1	0	1	0
N	0	-1	0	0	0	1	0	0	0	0	0	0
O	1	0	0	-1	0	-1	0	0	0	1	0	1
P	0	1	0	0	0	0	0	0	0	0	0	0
Margem de var.	1	3	1	1	1	2	1	1	1	1	1	1
Q	0	1	0	0	0	0	0	0	0	0	1	0
R	0	0	1	0	0	1	0	0	0	0	0	0
S	0	0	-1	0	0	-1	-1	0	0	0	0	0
T	0	-1	0	-1	0	0	-1	0	0	0	0	0
Margem de var.	0	2	2	1	0	2	1	0	0	0	1	0
Soma das margens	4	7	5	6	4	8	6	5	1	4	3	3

Controlos

As seguintes regras não têm nada de mágico. Para as formular, G. Peaker inspirou-se nas técnicas de controlo de qualidade utilizadas pela indústria. Experiências insistentes no domínio dos exames permitiram adaptá-las.

O controlo assim verificado é racional, rápido, eficaz, mas não muito apurado. Por conseguinte, se, apesar da aproximação grosseira dos critérios, um dos aspectos da correcção não parece aceitável, há sem dúvida um problema!

1.º Severidade

A experiência mostra que os esforços de harmonização devem incidir principalmente sobre esse ponto. Como vamos ver, em caso de problema, a solução felizmente é muito fácil.

Regra:

- a) Partir da média ¹ dos totais gerais (aqui 63).
- b) Não tolerar desvios para além duma margem central de 10 pontos ¹ em relação a essa média, ou seja 5 pontos para mais e 5 para menos (aqui 58-68).

Verificação:

Todos os totais gerais estão compreendidos nessa margem. Todos os examinadores são, portanto, duma severidade aceitável.

Solução em caso de excesso de severidade ou de generosidade: supunhamos que estamos perante os totais seguintes: 75-72-68-66-64-60-49-46-44-41-40.

Verifica-se que se reunirmos o mais severo e o mais generoso, depois o segundo mais severo e o segundo mais generoso, etc., as médias dos pares estão muito próximas das médias gerais.

Ou seja:			Média
(1)	40	75	57,5
(2)	41	72	56,5
(3)	44	68	56
(4)	46	66	56
(5)	49	65	57
(6)	60	74	62
Média geral			57,5
Margem de 10 pontos			52,5-62,5

Só os dois examinadores do par n.º 6 se situam entre 52,5 e 62,5. Estes dois serão autorizados a corrigir sozinhos.

¹ Os examinadores classificam 20 composições de 1 a 5. Dez pontos equivalem a uma diferença média de meio ponto por exercício para o conjunto dos 20 exercícios.

Nos outros casos, os exercícios deverão ser vistos por dois membros dum dos grupos constituídos e far-se-á a média dos pontos.

2.º *Discriminação*

Regra:

O total das margens médias (3) não pode ser superior ao dobro do total das margens dum examinador (4) e reciprocamente.

Verificação: Nenhum problema.

3.º *Conformidade*

Regra:

O total das margens das diferenças (5) entre as notas atribuídas por um moderador e a média das notas atribuídas pelo conjunto não deve ser superior a 12¹.

Verificação: Nenhum problema.

Conclusão

No caso presente, todos os moderadores foram bem sucedidos nas três provas de controlo. E assim poderão trabalhar sozinhos.

Em que situação nos encontramos? Das doze amostras de 20 exemplares, uma será corrigida agora definitivamente (foi vista por doze classificadores!)

Como cada um dos moderadores pode trabalhar sozinho, a sequência das operações será rápida.

D. Nova correcção das restantes amostras e controlo

As operações são praticamente as mesmas que para o controle dos moderadores. No entanto, damos um exemplo pormenorizado, porque a apresentação mais concisa (um único moderador e um único professor) permite uma melhor visão de conjunto.

¹ Mostramos na p. 232 que se obtém com este ponto de referência empírico uma boa avaliação da correlação entre duas séries de notas.

**CÁLCULOS PARA A COMPARAÇÃO
ENTRE UM PROFESSOR E UM MODERADOR
(1 é a melhor nota; 5 a menos boa!)**

5 grupos de 4	Alunos escolhidos ao acaso	Nota atribuída por		Diferença (Moderador-Professor)
		moderador	Professor	
1.º grupo	Henrique João Paulo Pedro	Mín. 5 4 Máx. 2 3	5 5 4 1 Máx. 1 2	Mín. 0 0 1 1 Máx.
	Margem Mín.-Máx.	3 14	12 4	Margem das diferenças = 1
2.º grupo	André Eduardo Júlio Renato	Máx. 2 1 1 Mín. 4	1 1 1 1 Mín. 4	Máx. 1 0 0 0 Mín.
	Margem Mín.-Máx.	3 8	7 3	Margem das diferenças = 1
3.º grupo	António Camilo Eugénio Jerónimo	Máx. 3 5 2 Mín. 6	3 3 5 5 3 3 Mín. 6	Máx. 0 0 -1 ¹ 0 Mín.
	Margem Mín.-Máx.	4 16	17 3	Margem das diferenças = 1
4.º grupo	Tiago Lourenço Martinho Vitor	Mín. 5 Máx. 1 4 4	5 5 1 1 3 3 3 3	Mín. 0 0 1 1 Máx.
	Margem: Mín.-Máx.	4 14	12 4	Margem das diferenças = 1
5.º grupo	Bruno Hugo Leão Simão	Máx. 3 Mín. 4 4 3	2 2 2 2 5 5 Mín. 2	Máx. 1 2 -1 1 Mín.
	Margem: Mín.-Máx.	1 14	11 3	Margem das diferenças = 3
Os cinco grupos reunidos	Total dos graus Margens reunidas	66 (1) 15 (3)	59 (2) (4) 17	Total das margens de diferenças = 7 (5)

Verificações

1.º Severidade dos classificadores

A diferença entre total (1) e total (2) não deve ser superior a 10.

Neste caso, $66 - 59 = 7$: aceitável.

2.º Discriminação

O total (3) não pode ser superior ao dobro do total (4) e reciprocamente.

No caso, 15 e 17: aceitável.

3.º Conformidade

O total das margens das diferenças (5) não deve ser superior a 12.

Neste caso, $= 7$: aceitável¹.

4.º Conclusão

No caso, não parece pôr-se qualquer problema. As notas do professor são aceites e as operações terminadas.

Se as notas do professor não parecerem aceitáveis, o ideal é provocar a discussão com ele a fim de encontrar um acordo.

Não é possível sempre, particularmente por falta de tempo.

A solução seguinte é mais rápida. Se um moderador encontrar um problema, convidará quatro dos seus colegas a corrigir de novo independentemente a mesma amostra de vinte exemplares. Calcula-se a média entre os cinco moderadores, e o professor é convidado a ajustar todas as notas em função das notas dos moderadores.

¹ Esta maneira simples de proceder evita o longo cálculo da correlação entre as duas séries de notas.

Obtém-se deste modo um coeficiente de correlação muito próximo do coeficiente calculado pelo método clássico, fazendo a operação:

$$\frac{152 + 172 - 72}{2 \times 15 \times 17} = .90$$

Segundo G. Peaker, os casos que exigem semelhante correlação tornam-se cada vez mais raros, à medida que o sistema de moderação funciona, em virtude das trocas de pontos de vista frutuozos que provocam em cada caso.

E. Como ajustar notas discordantes?

Trata-se quer dum simples ajustamento da média (ou da mediana) por subida ou descida das notas de todos os alunos (severidade ou benevolência excessivas), ou duma modificação da distribuição geral, para melhorar a discriminação.

1.º Ajustamento da mediana

Para um mesmo exame, é a seguinte a situação:

	<i>Quartil superior</i>	<i>Mediana</i>	<i>Quartil inferior</i>	<i>Desvio-padrão aproximado</i>
Moderador	16	13	11	$\sigma = 4$
Professor	11	8	6	$\sigma = 4$

O professor é mais severo que o moderador, mas também discrimina como deve ser. Se o professor junta 5 pontos a cada aluno, restabelece-se o paralelismo.

2.º Ajustamento da mediana e do desvio-padrão

Situação inicial:

	<i>Quartil superior</i>	<i>Mediana</i>	<i>Quartil inferior</i>	<i>Desvio-padrão aproximado</i>
Moderador	16	13	11	$\sigma = 4$
Professor	16	12	9	$\sigma = 5$

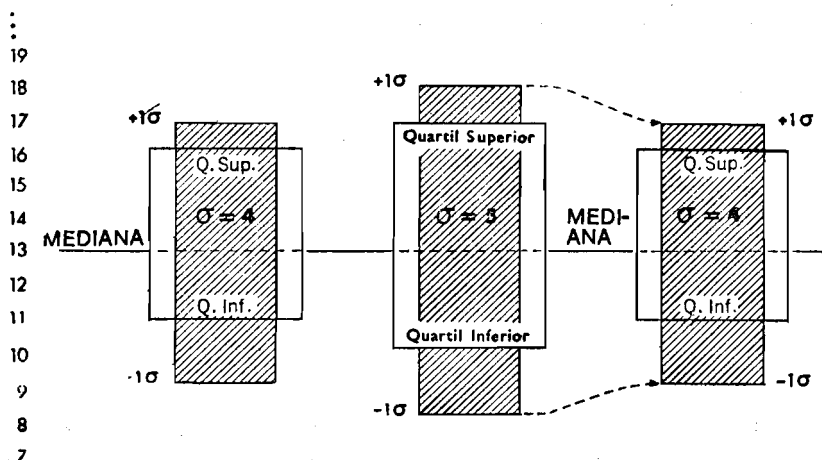
¹ Adaptado de D. MATHER et al., o. c., p. 183.

Para alinhar pelo moderador, corrige-se primeiro pela mediana; junta-se um ponto a todos os alunos.

Suponhamos que as notas iniciais do professor são:

... 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 ...
 |
 Mediana

Vamos mostrar primeiramente, mediante o desenho, em que consiste o ajustamento do desvio-padrão.



I. Notas do moderador

II. Notas do professor, aumentadas de um ponto. A mediana tornou-se igual à do moderador. Falta ajustar σ .

III. Notas do professor diminuídas uma segunda vez para alinhar o desvio-padrão pelo do moderador.

Para a operação III, o ajustamento em função do novo desvio-padrão é um pouco mais complicado que para a mediana. A construção duma tábua de conversão facilitará as operações.

Para achar a nova nota correspondente a $+1\sigma$, basta juntar o novo sigma à nova mediana, ou seja $13 + 4 = 17$. A nova nota correspondente a $-1\sigma = 13 - 4 = 9$, etc.

Para as notas que correspondem a frações de sigma, calcula-se primeiro o desvio em relação à mediana, multiplica-se por $4/5$ e soma-se o resultado à mediana. O seguinte exemplo vai esclarecer a situação.

	<i>Notas iniciais do professor</i>	<i>Notas ajustadas em função da nova mediana</i>	<i>Segundo ajustamento em função do novo Desvio-padrão (σ)</i>
	.	.	.
	19	.	.
	18	.	.
$+ 1 \sigma$	17	18	17 *
	16	17	16 **
	15	16	15 ***
	14	15	15 ****
	13	14	14
Mediana	12	13	13
	11	12	12
	10	11	11
	9	10	11
	8	9	10
$- 1 \sigma$	7	8	9
	6	.	.
	5	.	.
	.	.	.
	.	.	.

Como foram encontrados estes números?

* Nova mediana: 13

$$13 + 4 = 17$$

Novo desvio-padrão: 4

** Na segunda coluna, 17 fica a 4 pontos da mediana.

Multiplicar 4 por $\frac{4}{5} = 3,2$

Mediana $+ 3,2 = 16,2$, arredondado para 16.

$$\text{*** } 16 - 13 = 3,3 \times \frac{4}{5} = 2,4; \quad 13 + 2,4 = 15,4, \text{ arredondado para } 15.$$

$$\text{**** } 15 - 13 = 2,2 \times \frac{4}{5} = 1,6; \quad 13 + 1,6 = 14,6 \text{ arredondado para } 15.$$

F. Nota do fim do ano

Trabalho do ano + trabalhos práticos + teste

Problema ¹

Pretende-se classificar os alunos em grupos diferenciados no fim dos seus estudos. *A* será o grau superior e valerá 1, *E* será o grau inferior e valerá 5.

Resolveu-se ter em conta três elementos, avaliados todos eles segundo as mesmas escalas de cinco graus:

— trabalho do ano = T.A.

— trabalhos práticos = T.P.

— teste regional = T.R.

Ao teste regional atribui-se a mesma importância que aos dois outros elementos. Daí a ponderação:

$$\text{T.A.} = \frac{1}{4}$$

$$\text{T.P.} = \frac{1}{4}$$

$$\text{T.R.} = \frac{1}{2}$$

¹ Segundo D. MATHER et al., o. c., pp. 149-154.

Quadro geral dos graus e quadro dos resultados depois de ponderação e ajustamento final.

ATRIBUIÇÃO DO GRAU FINAL

Aluno n.º	Teste regional T. R.		Avaliação por escola		Ponderação			Total	Grau final
	Score	Nota	T.A.	T.P.	T.R. $\times \frac{1}{2}$	T.A. $\times \frac{1}{4}$	T.P. $\times \frac{1}{4}$		
1	59	2	3	3	1	1	$\frac{3}{4}$	$2\frac{3}{4}$	3
2	77	1	1	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	1	1
3	51	3	4	3	$1\frac{1}{2}$	1	$\frac{3}{4}$	$3\frac{1}{4}$	3
4	12	6	5	5	3	$1\frac{1}{4}$	$1\frac{1}{4}$	$5\frac{1}{2}$	6
5	53	3	3	4	$1\frac{1}{2}$	1	1	$3\frac{1}{2}$	4
6	40	4	2	5	2	$\frac{3}{4}$	$1\frac{1}{4}$	4	4
7	66	1	2	2	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$1\frac{1}{2}$	1
8	60	2	3	4	1	$\frac{3}{4}$	1	$2\frac{3}{4}$	3
9	38	5	4	5	$2\frac{1}{2}$	$1\frac{1}{4}$	$1\frac{1}{4}$	5	5
10	70	1	2	2	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$1\frac{1}{2}$	1
11	56	2	2	3	1	$\frac{1}{2}$	$\frac{3}{4}$	$2\frac{1}{4}$	2
12	69	1	2	2	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$1\frac{1}{2}$	1
13	44	4	4	5	2	1	$1\frac{1}{4}$	$4\frac{1}{4}$	4
14	64	2	1	2	1	$\frac{1}{4}$	$\frac{1}{2}$	$1\frac{3}{4}$	2
15	19	6	5	5	3	$1\frac{1}{4}$	$1\frac{1}{4}$	$5\frac{1}{2}$	6
16	49	3	2	4	$1\frac{1}{2}$	$\frac{3}{4}$	1	$3\frac{1}{4}$	3
17	54	3	2	1	$1\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$2\frac{1}{4}$	2
18	47	4	1	2	2	$\frac{1}{4}$	$\frac{1}{2}$	$2\frac{3}{4}$	3
19	52	3	3	1	$1\frac{1}{2}$	$\frac{3}{4}$	$\frac{1}{4}$	$2\frac{1}{2}$	2
20	48	3	3	3	$1\frac{1}{2}$	$\frac{3}{4}$	$\frac{3}{4}$	3	3
21	50	3	1	2	$1\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	$2\frac{1}{4}$	2
22	24	6	3	4	3	$\frac{3}{4}$	1	$4\frac{3}{4}$	5
23	61	2	4	2	1	1	$\frac{1}{2}$	$2\frac{1}{2}$	2
24	57	2	1	3	1	$\frac{1}{4}$	$\frac{3}{4}$	2	2
25	42	4	3	5	2	$\frac{3}{4}$	$1\frac{1}{4}$	4	4
26	35	5	5	2	$2\frac{1}{2}$	$1\frac{1}{4}$	$\frac{3}{4}$	$4\frac{1}{2}$	4
27	45	4	4	1	2	1	$\frac{1}{4}$	$3\frac{1}{4}$	3
28	41	4	2	4	2	$\frac{3}{4}$	1	$3\frac{3}{4}$	4
29	27	5	5	5	$2\frac{1}{2}$	$1\frac{1}{4}$	$1\frac{1}{4}$	5	5
30	43	4	3	2	2	1	$\frac{1}{2}$	$3\frac{1}{2}$	4
31	67	1	1	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	1	1
32	31	5	5	3	$2\frac{1}{2}$	$1\frac{1}{4}$	1	$4\frac{3}{4}$	5
33	72	1	2	2	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$1\frac{1}{2}$	1

As notas ponderadas em itálico serão ajustadas em função do teste regional (T.R.).

A leitura do quadro mostra-nos que o *aluno* 1 obtém as três notas ponderadas: 1,1 e $\frac{3}{4}$. Ou seja $2\frac{3}{4}$ no total; grau final 3*. O *aluno* 3: $1\frac{1}{2} + 1 + \frac{3}{4} = 3\frac{1}{4}$; grau final 3.

Qual a razão dum último ajustamento ao atribuir o grau final? Porque a adição dos graus provoca uma nova concentração em volta da média; a redução é tal que, numa escala final de 5 graus, toda a população poderá vir a encontrar-se no meio.

O exemplo fictício seguinte ilustrará o fenómeno.

ALUNOS	GRAU		
	Teste R.	Trabalho ano	Média
Pedro	1	5	3
Paulo	2	4	3
João	3	3	3
Rogério	4	2	3
Henrique	5	1	3
Dispersão	4	4	0

Se nos reportarmos ao quadro da p. 237, notamos que, entre os graus ponderados, se encontra um grande número de 1 (17 ao todo), ao passo que não se encontram mais que dois na coluna *Total*. As melhores notas ficaram afogadas na adição dos graus.

O processo seguinte garante uma melhor justiça distributiva entre os alunos duma mesma classe e um melhor alinhamento pelo nível regional, reconhecido como a norma mais segura.

* Os números em *italico* no quadro indicam que o moderador, com o acordo da escola, ajustou a nota do T. A. para a harmonizar com o nível médio da região.

Parte-se do quadro seguinte:

TESTE REGIONAL		TOTAL DOS GRAUS PONDERADOS		GRAU FINAL	
Grau (1)	N. de alunos que obtiveram esse grau (2)	Grau (3)	N. de alunos que obtiveram esse grau (4)	N. de alunos que foi atribuído Grau (5)	Grau (9)
1		1			1
		1 ¹ / ₄			
		1 ¹ / ₂			
		1 ³ / ₄			
2		2			2
		2 ¹ / ₄			
		2 ¹ / ₂			
		2 ³ / ₄			
3		3			3
		3 ¹ / ₄			
		3 ¹ / ₂			
		3 ³ / ₄			
4		4			4
		4 ¹ / ₄			
		4 ¹ / ₂			
		4 ³ / ₄			
5		5			5
Para além	Para além				Não classificado

Operações

1. Indicar na coluna (2) o número de alunos e calcular o total.
2. Começar a anotar na coluna (4) e parar quando o total iguala o da coluna (2). Neste momento, faz-se o primeiro corte. Para não prejudicar certos alunos, ultrapassa-se, de

acordo com as necessidades, o número da coluna (3) por forma a atingir o nível em que se parou.

Ver-se-á, por exemplo, no quadro que segue, que o grau 2 se atribui a 7 alunos e não a 6, porque 2 deles obtiveram $2\frac{1}{2}$. Não se podia atribuir evidentemente o grau 2 a um e o grau 3 a outro.

GRAU REGIONAL		TOTAL DOS GRAUS PONDERADOS		GRAU FINAL	
Grau (1)	N. de alunos que obtiveram esse grau (2)	Grau (3)	N. de alunos que obtiveram esse grau (4)	N. de alunos a que foi atribuído (5)	Grau (6)
1	JHT I 6	1	// 2	6	1
		1 1/4	/// 4		
		1 1/2			
		1 3/4	/ 1		
2	JHT I 6	2	/ 1	7	2
		2 1/4	/// 3		
		2 1/2	// 2		
		2 3/4	/// 3		
3	JHT II 7	3	/ 1	7	3
		3 1/4	/// 3		
		3 1/2	// 2		
		3 3/4	/ 1		
4	JHT II 7	4	// 2	7	4
		4 1/4	/ 1		
		4 1/2	/ 1		
		4 3/4	// 2		
5	IIII 4	5	// 2	4	5
Para além	III 3	Para além		2	Não classificados

Conclusão

Nenhum dos sistemas descritos é perfeito, mas todos permitem melhorar o sistema tradicional de classificação.

Vai operar-se uma primeira escolha em função do fim em vista: selecção ou comparação. Foi sobretudo este segundo aspecto que reteve a nossa atenção.

Não nos pronunciamos em favor dum sistema particular. A decisão pertence às autoridades políticas e pedagógicas e aos docentes.

Deve fazer-se a todos os níveis um intenso esforço de experimentação: pequenos grupos de professores, investigações regionais e nacionais em colaboração com os centros psico-médico-sociais e com os laboratórios de pedagogia experimental das universidades.

À medida que os trabalhos forem avançando, ver-se-á provavelmente surgir um novo sistema que tenha em conta as nossas tradições bem como certos objectivos particulares do nosso país (a Bélgica). Faz-se sentir de novo, e quase imediatamente, a necessidade de activos centros de investigação pedagógica regionais. Voltaremos a este assunto.

QUINTA PARTE

UMA PEDAGOGIA DA DOCÊNCIA

(MASTERY LEARNING)

O PERIGOSO MITO DA CURVA DE GAUSS

Nas ciências humanas, a curva de Gauss em forma de sino desempenha um papel considerável, porque é a própria imagem da distribuição de muitas aptidões e qualidades. Os indivíduos médios abundam, mas os génios e os idiotas, os gigantes e os anões são raros.

A curva de Gauss é ou o reflexo da lei do acaso que preside ao nosso nascimento, ou a resultante da influência dum grande número de factores que agem duma maneira mais ou menos independente sobre um indivíduo ou sobre um objecto.

Como os testes medem muitas vezes aptidões, traços de personalidade ou capacidades de vastas populações, é natural que sejam aferidos segundo a distribuição gaussiana; aproximadamente 70% de médios, 13% de bons, 13% de medíocres 2% de muito bons, 2% de muito maus.

No decurso da construção desses testes, eliminam-se particularmente as perguntas que se viriam a revelar fáceis para muitos ou muito poucos indivíduos. O fim em vista é classificar cada um, atribuir-lhe o lugar que lhe pertence num grupo numeroso. Numa palavra, trata-se de organizar uma espécie de concurso, em que o melhor virá a ocupar necessariamente o primeiro lugar.

É por isso que muitos testes de aptidões ou inventários de conhecimentos são óptimos *instrumentos de selecção*.

O professor, na sua turma, tem um objectivo totalmente diferente. O ideal não é que *todos* os alunos aprendam a ler, a calcular e, dum modo geral, a dominar perfeitamente todos os conhecimentos julgados necessários ou úteis pela sociedade. *Instruir não é seleccionar*. Pelo contrário! É procurar que *todos* consigam. *Por conseguinte é lutar contra a curva de Gauss* tomada como modelo de selecção.

As consequências pedagógicas dessas observações são particularmente importantes.

CAPÍTULO 1

EVOLUÇÃO DA CURVA DOS CONHECIMENTOS

Quando, no nosso sistema de classes rígidas, um professor recebe, no dia do começo das aulas, um grupo de alunos que não conhece, encontra-se normalmente perante... duas curvas: uma representa a distribuição das aptidões e a outra a dos conhecimentos.

I. Curva das aptidões

Na acepção corrente — que provisoriamente mantemos — a palavra *aptidão* designa características inatas ou adquiridas, considerada como sintomáticas da capacidade dum indivíduo para adquirir um nível de competência mais ou menos elevado, em determinado domínio.

No ensino não estritamente especializado (que ultrapassa os vinte anos para grande parte dos nossos alunos) a própria amplitude do quadro de conhecimentos e capacidades a fazer adquirir (Matemáticas, Línguas, Ciências Naturais, Artes, etc., ...) torna impossível a selecção muito rigorosa, segundo uma aptidão particular.

Por isso, até um nível muito avançado da escolaridade, as aptidões dos alunos permanecem distribuídas ao acaso.

Em tais condições, um professor de matemática, por exemplo, que medisse as aptidões especiais pela sua disciplina obteria normalmente uma distribuição gaussiana ou — dado o

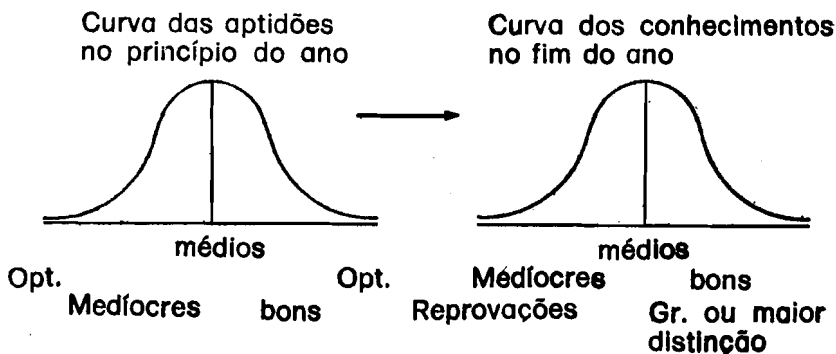
número provavelmente pouco elevado de alunos — um simples esboço dessa distribuição.

Os professores, aliás, pouca necessidade têm de recorrer a testes para conhecer a situação. Sabem por experiência que os médios constituirão a maioria e que os ótimos serão muito raros...

No entanto, produz-se geralmente nos espíritos uma distorção singular. Considera-se esta distribuição de aptidões como prognóstico de resultados escolares do fim do ano e fixa-se o nível de ensino de forma a que o prognóstico se verifique: será «medianamente» difícil, dando mesmo assim aos melhores a possibilidade de se revelarem ao passo que limita extraordinariamente a possibilidade dos «mediocres».

Temos de insistir em que a dificuldade «média» é determinada pela média das aptidões do grupo considerado e não por uma média de dificuldade *objectiva* das noções a ensinar ¹

Nessas condições, se o professor faz o mesmo curso a toda a classe, é normal que a *curva dos conhecimentos* adquiridos no fim do ano respeite também a distribuição gaussiana.



Será deste modo respeitada a vocação do ensino?

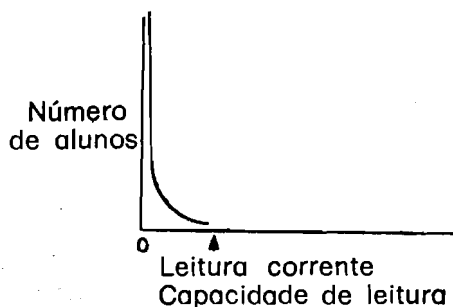
¹ Assim se explicam as diferenças consideráveis do nível médio segundo as classes e regiões, de que falámos na primeira parte desta obra.

II. Curva de conhecimentos

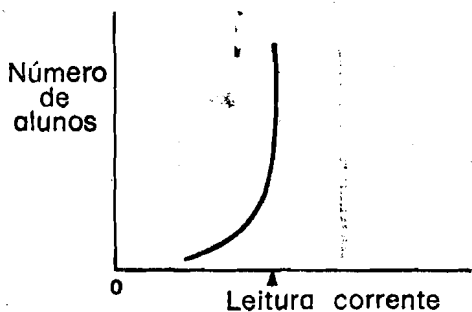
Voltemos ao primeiro dia do ano escolar e, em vez de considerarmos as aptidões específicas dos alunos, examinemos os seus conhecimentos.

Logicamente, o papel do professor é suscitar a aprendizagem de conhecimentos *novos*. Claro que considerar todos os alunos que formam uma turma como possuindo exactamente a mesma quantidade e qualidade de conhecimentos, é utópico. No entanto, o sistema de classes fixas que nós praticamos assenta na hipótese de que *todos* se encontram ao mesmo nível aproximadamente. Caso contrário, como ousaríamos nós ministrar para todos o mesmo tipo de ensino?

E na realidade? Consideremos o caso da entrada na primária. A maior parte das crianças não sabe ler. Alguns estão no bom caminho. Dois ou três lêem já correntemente. Neste momento, a curva da capacidade em leitura toma aproximadamente a forma da letra *i*.

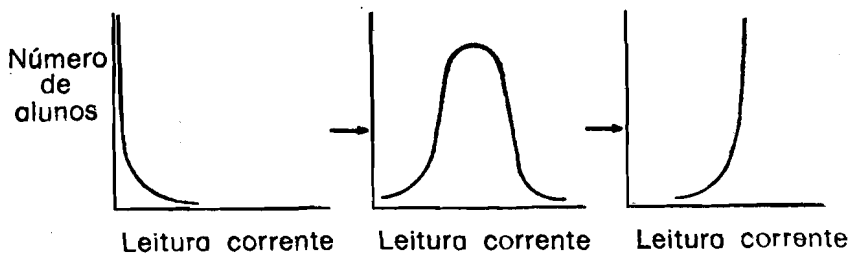


Embora a curva das aptidões específicas para a leitura seja muito provavelmente gaussiana, o professor não admite logo que só uma parte da classe possa aprender a ler. Antes pelo contrário. Contanto que não se caia no exagero da perfeição, podemos dizer que, depois dum ou dois anos, a maior parte dos alunos saberá ler correntemente. A curva dos conhecimentos terá mudado completamente de forma. Assemelhar-se-á a um *j*.



Entre a curva em i e a curva em j , é provável que tenha existido um momento em que os mais dotados tenham avançado mais depressa, os menos dotados se tenham atrasado e os médios se tenham situado entre ambos.

A evolução é, pois, esquematicamente a seguinte:



Capacidade de leitura

No caso da leitura, não se concebe que os pais ou responsáveis pelo ensino possam contentar-se com uma evolução diferente.

Mas, a partir de que grau de escolaridade, deixará de ser aceitável esta exigência?

CAPÍTULO 2

UMA PEDAGOGIA DE CURVA EM J

Um exame em que os resultados se distribuem segundo a curva em sino de Gauss presta-se bem à selecção, ao concurso.

Ora, por três razões históricas principais, o espírito de concurso impregnou o nosso ensino durante séculos.

- 1.º Por razões socio-económicas, apenas uma parte da população escolarizável tinha acesso à escola. Há cem anos, as famílias modestas faziam a primeira selecção não deixando acabar a escola primária a não ser aos filhos mais dotados, isto é, aos que obtinham melhores resultados nos seus exames. As bolsas de estudo, parcimoniosamente distribuídas, obtinham-se mediante concurso.
- 2.º Até há bem pouco tempo, — antes da era do ordenador — todos os países industrializados dispunham dum enorme excesso de matéria cinzenta. Explorava-se a que se apresentava mais barata e com mais facilidade, isto é, procurava-se sobretudo identificar os mais dotados. Mesmo para as crianças das famílias afortunadas, o ensino secundário desempenhava um papel selectivo.
- 3.º Os conhecimentos psicológicos e pedagógicos eram rudimentares. Os professores não estavam à altura de aplicar tratamentos delicados aos alunos com dificuldades de aprendizagem.

Hoje ainda fazem repetir muitas vezes um ano ao aluno que não conseguiu obter os resultados mínimos exigidos para a passagem. Por outras palavras, em vez de aplicarem

remédios particulares, contentamo-nos com colocar de novo o aluno nas condições (mesmo professor, mesmo método) em que o fracasso se produziu.

Sendo cada ano escolar considerado como um filtro para o seguinte, parecia normal um certo número de fracassos. Em resumo, a repartição gaussiana parecia satisfatória.

Todavia, a preocupação constante da promoção dos mais aptos causou progressivamente uma deformação pedagógica mais subtil. Chegou-se a pensar que os conhecimentos «que verdadeiramente contam», os conhecimentos «profundos», o «verdadeiro jogo das ideias abstractas», só são assimiláveis por uma minoria possuidora de aptidões específicas num grau elevado. E, em função disso, forjaram-se os métodos de ensino.

Os alunos e os pais, aliás, aceitam este estado de coisas sem grande discussão. Raramente se inscrevem na secção «latim-matemática» sem se sentirem especialmente dotados nesse domínio.

Evidentemente que não podemos ambicionar fazer indiferentemente de cada um *virtuoso* da matemática, do piano ou da literatura. Onde se situa porém, o limite? A partir de que grau é impossível o acesso a um determinando saber para os que possuem aptidões médias ou até medíocres?

Por natureza, cada um de nós é mediocrementemente dotado em muitos domínios. Não faltam, por exemplo, intelectuais mais ou menos refractários à matemática. Todavia, observa-se frequentemente que, à força de vontade, de perseverança, de insistência na repetição das explicações, de mudança de manual ou de professor para encontrar uma forma de ensino que convenha, certas noções de matemática, primeiro consideradas inacessíveis, são lindamente dominadas por determinadas pessoas.

Contanto que consagrem a isso o tempo necessário, os alunos médios (isto é, tendo em conta a selecção operada nos estudos anteriores, provavelmente mais de 80% da população escolar dum determinado nível) podem ir muito mais além do que se imagina.

Dentro desta perspectiva, a nova definição da aptidão proposta por J. Carroll adquire todo o seu valor. «A aptidão é a quantidade de tempo exigida pelo aprendiz para dominar uma matéria»¹.

As implicações desta concepção são consideráveis. Se lhes concedermos o tempo necessário, todos os alunos que fazem normalmente parte duma turma podem atingir um bom ou mesmo muito bom resultado.

¹ J. CARROL, A Model of School Learning, *Teachers College Record*, 1963, 64: 723-733.

CAPÍTULO 3

A TEORIA DA AVALIAÇÃO FORMATIVA

A expressão *avaliação formativa* — oposta a *avaliação somativa* (tratada no capítulo IV) — foi forjada por Michael Scriven¹.

A *avaliação normativa* é-nos agora familiar. Para interpretar o *score* obtido com um teste clássico de inventário de conhecimentos ou de inteligência, situá-lo-emos numa distribuição estatística. O resultado obtido por um indivíduo é julgado em função de outrem. Da mesma maneira classificam-se ainda muitas vezes os alunos entre si segundo a ordem crescente ou decrescente dos seus resultados escolares e é pela posição ocupada que muitos pais avaliam o trabalho dos filhos.

Em ambos os casos, porém, o resultado é essencialmente relativo. Se o grupo de referência variar de composição, o resultado aparecerá a uma outra luz.

Uma simples diferença de idade também pode mudar consideravelmente a face das coisas. Em muitas normas de testes de conhecimentos de uso corrente, um ano de idade ou um ano lectivo a mais ou a menos bastam para que um mesmo resultado seja considerado medíocre ou bom.

¹ M. SCRIVEN, *The Methodology of Evaluation*, in R. STAKER, (Ed.), *Perspectives of Curriculum Evaluation*, Chicago, Rand McNally, 1967.

Ver ainda B. S. BLOOM, *Learning for Mastery*, in *Evaluation Comment*, 1968, 2. A nossa exposição do método de avaliação formativa apoia-se directamente nessa publicação. Numerosos resultados de experiências em que a teoria do *mastery learning* foi aplicada, foram publicados por J. BLOCK, B. BLOOM, HASTINGS (1970).

Que se tenha ou não adquirido o conhecimento não foi por conseguinte a preocupação dominante dos construtores desses testes mas antes a que cadência essa aquisição se realizou. Reportando-nos à definição de J. Carroll, avalia-se a aptidão em vez de avaliar o conteúdo da aprendizagem.

Os defensores da *avaliação formativa* tomam uma posição oposta.

Visto que o aluno vem à escola para aprender, o que importa não será situá-lo na ascensão dos conhecimentos?

Imaginemos que uma análise científica rigorosa revele que, com todas as cambiantes e complicações, a concordância do participio passado, empregue em francês com o auxiliar ter, apresente cem casos diferentes. Para cada um deles, podemos definir critérios de domínio;

Pode então fixar-se o número de casos a dominar, segundo o nível escolar. Neste contexto, muda de natureza a avaliação escolar.

O aluno é classificado em função dum critério objectivo: o caminho percorrido na aquisição.

A avaliação formativa consiste pois fundamentalmente em dividir uma tarefa, um curso, uma matéria,... em *unidades* e em determinar para cada unidade em que medida um aluno domina a dificuldade. Trata-se duma intenção diagnóstica.

A unidade neste contexto é uma porção precisa duma aprendizagem a efectuar. Muitas vezes define-se sob a forma de objectivo ou até de micro-objectivo a atingir. As unidades podem ser hierarquizadas entre si, na medida em que o domínio duma é necessário para abordar outra ou outras.

No domínio duma dada unidade, podemos igualmente distinguir uma hierarquia de processos mentais (por exemplo, segundo a taxonomia de Bloom).

Seja como for, a *avaliação formativa* tem por único fim reconhecer onde e em que o aluno sente dificuldade e procurar

informá-lo. Esta avaliação não se traduz em nota, nem muito menos em *scores*. Trata-se dum *feedback* para o aluno e para o professor.

Em consequência da sua *natureza diagnóstica*, a avaliação formativa exige a acção correctiva, sem a qual não existe verdadeiro ensino.

Além disso, se desejarmos conduzir todos os alunos até ao nível de conhecimentos mínimo, ou até mesmo ao domínio total da noção ou da técnica, a própria maneira de ensinar será posta em causa. Não será possível aplicar indiferentemente a todos um único e mesmo método durante um mesmo período de tempo.

O que viria a ser o ensino, segundo essa perspectiva?

Põem-se três problemas cruciais:

- 1.º Como escalonar a ascensão do saber?
- 2.º Como conduzir o aluno?
- 3.º Onde se situa o limite prático dessa pedagogia da curva em *j*?

Que eu saiba, não existem ainda respostas completas para essas perguntas. Uma tal imprecisão do saber parece normal em ciências naturais e em medicina. Deveremos habituar-nos a aceitá-la igualmente nas ciências da educação. A nossa discussão acaba muitas vezes em investigações a emprender ou a continuar.

I. Escalonar a ascensão do saber

Em viagem, para determinar a que distância nos encontramos do fim, devemos obedecer a duas condições: por um lado saber onde estamos e para onde vamos e, por outro, dispor dum mapa que nos indique claramente o caminho.

Do mesmo modo, em educação, devemos definir os objectivos a atingir e determinar com precisão as aprendizagens particulares para os atingir. O problema varia segundo se trata ou

não de aprendizagens de base. Para estas últimas, teremos de descobrir o encadeamento «crítico» das matérias, ou seja o ponto em que uma delas não é acessível se a antecedente não tiver sido assimilada. Em virtude disso, a aprendizagem dos conhecimentos e das técnicas de base, em última análise, deve ser linear, ao passo que as aquisições e as aplicações que vão para além vêm abrir-se diante delas caminhos cada vez mais numerosos.

Por exemplo, qualquer que seja o método de ensino, não é possível aplicar completamente uma regra de três sem conhecer — especificamente — a noção de multiplicação e de divisão. Num nível mais elevado, como fazer cálculo integral sem saber o que é uma função? Semelhantes afirmações são ainda muito vagas. Quais são exactamente as aprendizagens necessárias e suficientes para poder assimilar a regra de três? E, dentre elas, quais são críticas em relação às outras?

A definição das unidades de aprendizagem põe um problema ainda longe de estar resolvido. Nos pontos extremos, afrontam-se duas escolas: uma puramente empírica, outra experimental.

Os empiristas referem-se à sua experiência profissional, à sua lógica e também à sua intuição, para dividirem uma matéria em partes relativamente homogêneas e para as ordenar. É assim que procedem os autores dos manuais quando dividem uma matéria em secções, capítulos, rubricas, parágrafos e alíneas.

Já vimos que o grupo de Popham adoptou um método empírico também, mas já mais rigoroso. Chega ao nível dos objectivos específicos.

Pelo contrário, os defensores da análise hierárquica dos conteúdos, principalmente R. Gagné e J. B. Carroll, procuram regras psicológicas ou psicométricas que permitem reconhecer a estrutura duma matéria e os momentos de aprendizagem obrigatórios.

Observa-se porém que raras são as vias únicas nas aprendizagens. Para a pedagogia correctiva, importa mais um conhe-

cimento claro das componentes, das unidades, pois que, no final de contas, todo o indivíduo deve gozar duma liberdade total para as estruturar, repartir, articular e chegar a um maior domínio funcional do meio.

Diremos mais simplesmente que pouco importam os caminhos escolhidos para aprender ou fazer aprender uma mesma matéria, desde que a psicologia da criança seja respeitada e que na ocasião desejada o professor saiba em que pontos fazer incidir o seu controlo para assegurar uma aprendizagem efectiva.

É evidente que certos ramos do saber se prestam melhor que outros a uma estrutura de unidades hierarquizadas. Em larga medida, prestam-se especialmente a isso as matemáticas. Ao contrário, a história universal, o ensino da composição, pouco assentam em estruturas hierarquizadas lógicas ou psicológicas. Os factos abundam neles mas as regras são raras. Não existe algoritmo operacional da redacção.

Por conseguinte, as aprendizagens nesses ramos estruturalmente fluidos, não são cumulativos mas aditivos. Mesmo até no interior duma unidade, as componentes não mantêm sempre relações estreitas e não se ordenam necessariamente segundo um modelo bem articulado. Sabe-se que, para a história por exemplo, se procura contornar a dificuldade, respeitando a ordem cronológica ou trabalhando por temas, mas isso não passa de artifícios metodológicos.

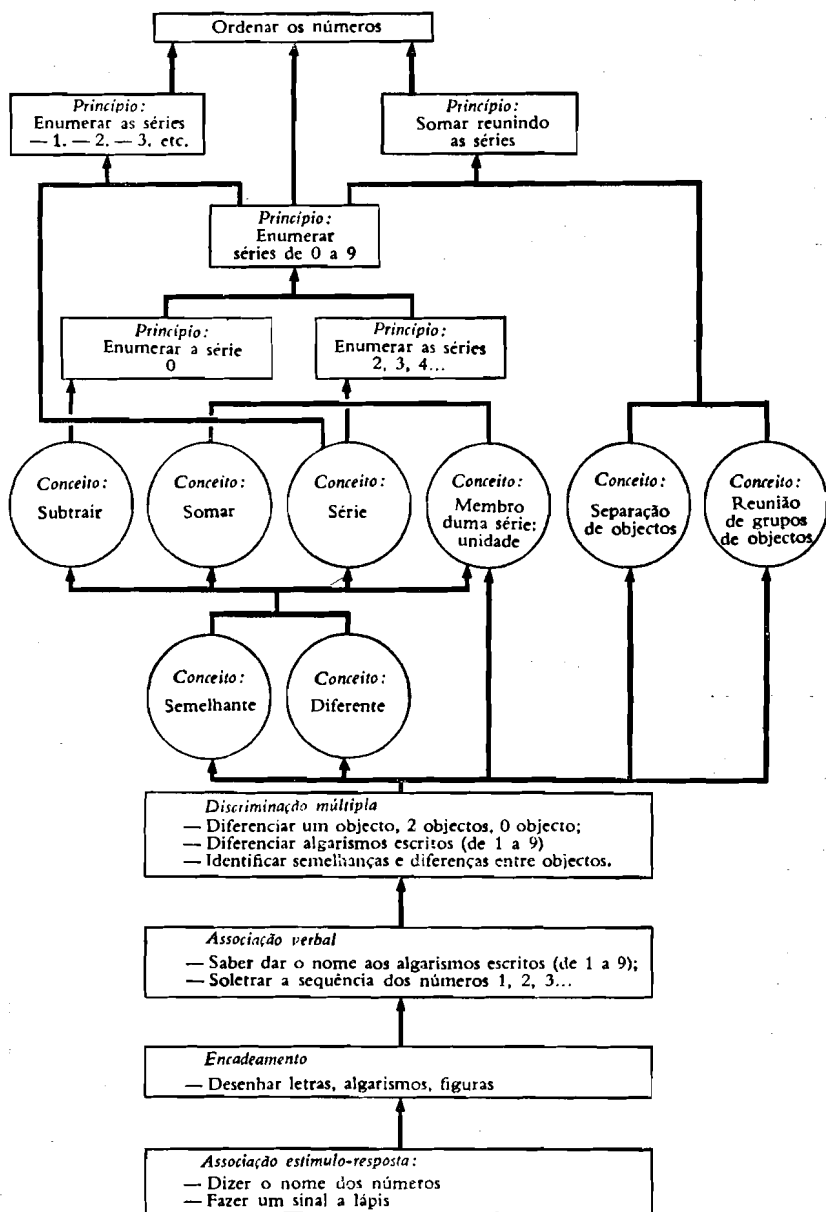
Teremos de concluir que as unidades e portanto a avaliação formativa não terão lugar nestas matérias? Com certeza que não, pois de outra maneira o ensino teria de abandonar-se a uma improvisação permanente de que é fácil imaginar os inconvenientes e as lacunas.

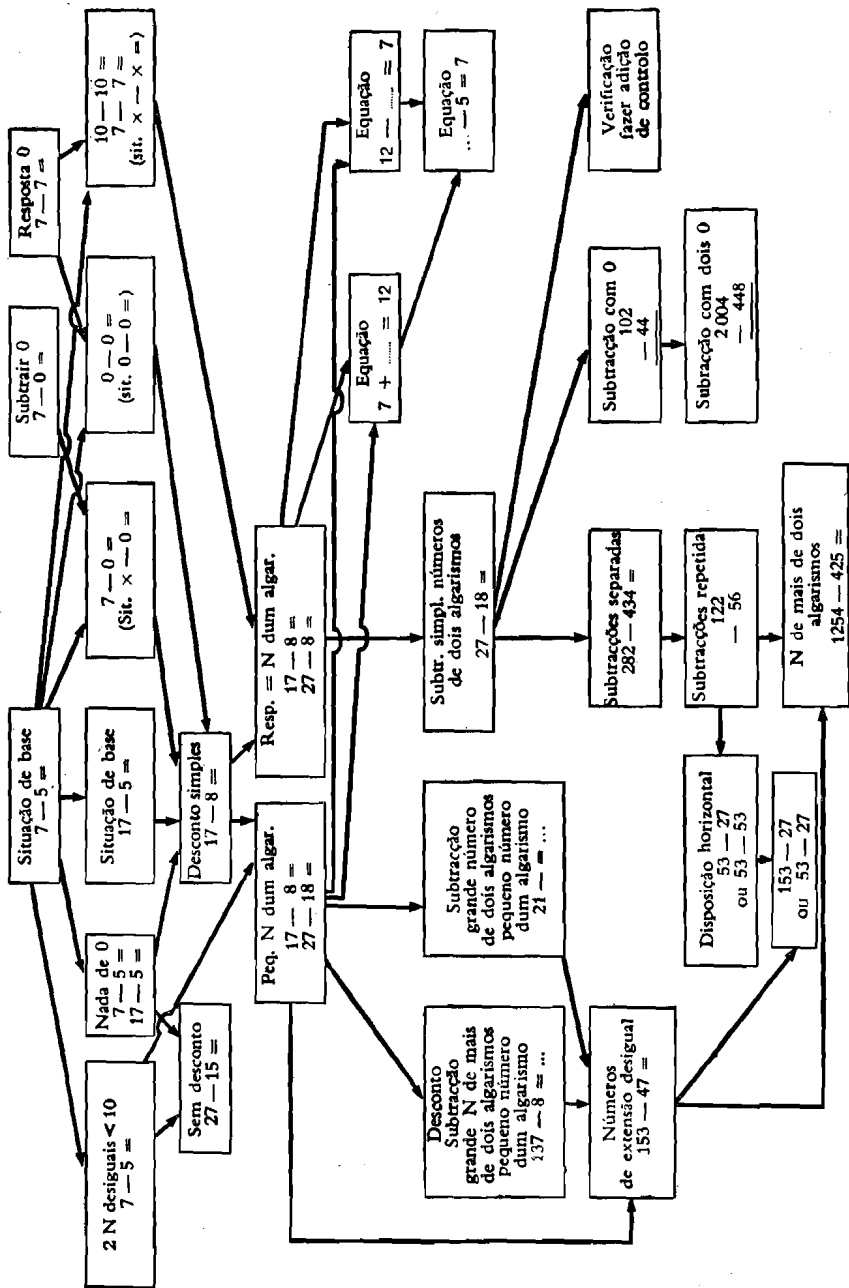
Fora dos imperativos psicológicos e axiológicos, o educador fica livre de ordenar as aprendizagens a realizar *como se pudessem hierarquizar-se*. A ordem poderá ser corrigida de acordo com as necessidades pedagógicas.

Deste modo, dispõe-se ao menos duma linha de conduta e duma possibilidade de controlo docente. Os objectivos centrados sobre os princípios, as estruturas e os *transferts* terão naturalmente a prioridade. Os outros, mesmo se são de natureza mais atomística, merecem ser rigorosamente inventariados, quando mais não seja para evitar estéreis acumulações factuais.

Apresentamos dois exemplos de estrutura hierárquica de aprendizagem das operações numéricas¹. Uma, proposta por Gagné, é psicológica, outra, proposta por Hively, é lógica.

¹ Vamos buscar a adaptação francesa deste exemplo a D. BONORA, Les buts de l'éducation, in M. REUCLIN. *Traité de psychologie appliquée*, 6, o. c., p. 15.





II. Guiar o aluno

Uma vez definido claramente o caminho que conduz a uma aprendizagem, temos de empenhar nisso o aluno. Para as aquisições fundamentais, resolvemos ainda que, em princípio, *todos* deveriam chegar ao fim.

Neste caso, é essencial a avaliação frequente dos progressos. Daí a necessidade de *exercícios de domínio* e de *testes diagnósticos* que incidam sobre matérias muito limitadas e utilizados pelos próprios professores.

Estes instrumentos — que só a colaboração dos docentes e dos investigadores permitirá construir em número suficiente — indicam *em que ponto se encontra o aluno* (o que lhe dá a nota) e *onde sente dificuldades*.

Os *remédios* são múltiplos e devem entrar na nossa prática pedagógica quotidiana.

- a) Indicar de maneira precisa a parte do curso a estudar de novo.
- b) O trabalho em subgrupos: se um grupo de mais de três alunos encontra a mesma ou as mesmas dificuldades, o professor terá vantagem em consagrar uma parte do tempo da lição a esse grupo, trabalhando os outros independentemente.
- c) Se o grupo tem o máximo de três alunos, B. Bloom sugere que cada um deles seja integrado num grupo heterogéneo de dois ou três. Como os alunos deixam de estar em concorrência, mas lutam pelo domínio duma matéria, entra em acção a ajuda mútua.
- d) Nas escolas onde há vários professores duma mesma disciplina, o aluno deveria ter a faculdade de pedir uma explicação, até mesmo uma pequena lição particular a um outro professor diferente do seu.
- e) Variar a apresentação, pondo à disposição dos alunos manuais que explicam a noção de diversas maneiras.

Permitir o recurso ao ensino programado, aos meios audio-visuais. Variar o nível de abstracção.

- f) Dum modo geral, não impor a todos o mesmo tempo para as mesmas coisas.

Semelhante prática resulta. B. Bloom refere uma experiência conduzida nesse sentido em que mais de 80% dos alunos obtiveram a melhor nota no fim do ano. É o triunfo da pedagogia da curva em *j*.

Será sempre assim? Nada permite afirmá-lo.

III. A relação tempo-aprendizagem

À medida que a investigação avança, vai-se tornando precisa a relação entre o factor tempo e as aprendizagens. A visão de conjunto mais recente sobre este problema é-nos dada por B. Bloom¹. Encara sucessivamente a evolução das aprendizagens à medida que os anos civis passam, em função do tempo decorrido para dominar uma aprendizagem e em função do tempo durante o qual o aluno se dedica à aprendizagem.

À medida que os anos passam, vão-se realizando progressos importantes. Na escola primária, um nível de *performance* atingido por 20% dos alunos superiores num determinado ano, é atingido muitas vezes por 50% dos alunos, no ano seguinte, e por 80%, dois anos após.

Se considerarmos mais particularmente o tempo que decorre antes que os diferentes alunos dominem uma aprendizagem especial, constatamos, com uma aproximação grosseira, que os alunos lentos lhe dedicam cinco vezes mais tempo que os alunos rápidos. Esta relação de 1/5 verifica-se muito

¹ B. S. BLOOM, *Time and Learning*, Comunicação ao 81.º Congresso anual da *American Psychological Association*, 1973.

bem se utilizarmos os *scores* obtidos nos testes de aptidão e nos testes de inteligência geral para prever o tempo necessário a uma aprendizagem (correlação de +.50 a +.70).

No entanto, estes dados dizem respeito ao ensino tradicional. Que se irá passar se aplicarmos uma metodologia centrada no domínio das aprendizagens?

Por ocasião da aprendizagem da primeira unidade, verifica-se geralmente a relação 1/5. Recordemos que neste caso a diferença de tempo representa a ajuda suplementar dada directa ou indirectamente para suprir as lacunas reveladas pela avaliação formativa.

Se esta primeira unidade é exigida para a aprendizagem da segunda, e se o estudo desta for abordado a partir do momento em que a primeira for adquirida, a relação temporal de 1/5 tende a reduzir-se, para se situar à volta de 1/3 depois de uma série de unidades. Chegado a este ponto, B. Bloom faz uma observação capital: «...neste processo de aprendizagem centrado no domínio da matéria, o significado prognóstico dos testes de inteligência geral ou de aptidões, no que diz respeito ao tempo a consagrar às aprendizagens, diminui progressivamente para atingir por fim um valor muito baixo. Ao contrário, o *score* de um teste formativo é bom para a previsão do tempo necessário à aprendizagem da unidade que vem depois na sequência».

Por outras palavras, a taxa de *performance* dos alunos vai melhorando à medida que eles progridem no programa. O particular auxílio a dar-lhes para dominarem as unidades de aprendizagem vai diminuindo paralelamente.

Um terceiro modo de encarar as coisas é considerar apenas o tempo em que o aluno se dedica plenamente a uma aprendizagem determinada, consagrando-lhe atenção e energia (*time on task*). É evidente que uma hora de curso individualizado é diferentemente utilizada pelos alunos. Uns lançam-se imediatamente ao trabalho e não o abandonam; outros não começam logo e cansam-se depressa, etc.

Daqui a utilidade de fazer a distinção entre o tempo decorrido globalmente e o tempo concentrado no trabalho.

Importa evidentemente ter em conta comportamentos observáveis e comportamentos não observáveis (reflexão sobre o problema), avaliados por introspecção. Como seria de esperar, o tempo concentrado na aprendizagem é bom indicador da *performance* (após correcção para a fidelidade, as correlações dão conta de cerca de três quintos da variação das *performances* entre estudantes)¹.

Todavia, exerce aqui também uma influência espectacular uma pedagogia da curva em *j*. Consideremos dois grupos equivalentes de alunos. A um é ministrado o ensino habitual, o outro é sujeito ao sistema centrado no domínio dos objectivos.

Por ocasião da aprendizagem da primeira unidade de domínio da matéria, o tempo de concentração dos dois grupos pouco difere e eleva-se, em média, a 65% do tempo decorrido. À medida que as unidades se sucedem, observaremos que os alunos empenhados no processo de domínio da matéria se vão concentrando cada vez mais (até 85% do tempo decorrido), ao passo que o esforço dos outros se vai relaxando (até 50% do tempo decorrido). Por outras palavras, «o grupo centrado no domínio da matéria aprende mais eficazmente a aprender».

Outra observação capital. À medida que avançamos na aprendizagem centrada no domínio da matéria, vão-se reduzindo as diferenças entre estudantes. A princípio, as variações no tempo concentrado na tarefa a dominar são da ordem de 1 a 3. À medida que se avança, tende a reduzir-se até 1 e 1,5, por vezes menos. Trata-se pois dum nivelamento por cima na eficácia das aprendizagens.

Todas estas observações são ainda muito recentes e muito parcelares para permitirem uma generalização. Vemos, no entanto, as perspectivas extraordinárias que abrem à pedagogia

¹ B. BLOOM nota com razão que o tempo consagrado pelos alunos à aprendizagem pode também ser considerado como índice da qualidade do ensino.

e as esperanças que dão a tantos estudantes sujeitos muitas vezes a inferioridades duradoiras em virtude do seu meio social de origem ou duma escolaridade anterior pouco favoráveis às novas aprendizagens a realizar.

Como já teremos pressentido, o nivelamento por cima tem no entanto os seus limites. E teremos ainda verificado que, em nenhum dos números apresentados anteriormente, se pretende igualizar por completo. É também evidente que não tivemos sempre em conta populações escolares normais, oferecendo é certo uma larga margem de variação de aptidões à partida, mas excluindo inferioridades mentais graves. Claro que não é possível permitir que alguém passe vinte anos em vez de seis no ensino secundário!

Esta verificação não deve, contudo, pôr em causa o sistema. O fundamental é orientar progressivamente os alunos em função das suas aptidões reais. O avanço tão rápido quanto possível num ensino centrado sobre o domínio das aprendizagens seria sem dúvida um bom indicador da necessidade de reorientar ou pelo menos de proceder a um exame aprofundado da situação do aluno. Os ciclos de observação e de orientação, introduzidos progressivamente no ensino secundário da maior parte dos países, dão-nos essa possibilidade.

IV. O sistema de ensino posto em causa

A pedagogia da curva em *j* provoca o nivelamento pelo nível superior. Seria no entanto quimérico imaginar que ela permite o nivelamento pelo... génio.

Já vimos que menos de cinco por cento da população possui aptidões *excepcionais*. Estas devem ser desenvolvidas ao máximo, tanto por respeito pela pessoa como no interesse da comunidade.

Um sistema de ensino flexível permitirá instruir toda a gente a uma cadência apropriada e ao mesmo tempo deixar que desabrochem os superdotados.

Aos programas rígidos substituem-se actividades multiformes. Uma vez o trabalho é individualizado, em particular graças à tecnologia tradicional; outras vezes trabalha-se por grupos homogeneizados de acordo com as aptidões específicas para uma única matéria¹; noutros casos, constituem-se grupos de grandeza variável segundo critérios afectivos.

Nas actuais condições, a classe é um colete de ferro. Por que razão um aluno de seis anos que já sabe ler ao entrar na primeira classe da primária não poderá participar nos exercícios de leitura da segunda classe? Porque é que determinado aluno do 4.º ano do ensino secundário não haveria de poder seguir um curso de matemática do 5.º ou do 6.º?

Nas velhas escolas primárias da aldeia, em que às vezes um único professor se via a braços com seis anos, era normal fazer estas deslocções. Nas escolas secundárias que actualmente têm várias classes da mesma idade, é possível trabalhar a quatro ou cinco níveis de aptidões diferentes, pelo menos nalgumas disciplinas principais. Um sistema destes funciona perfeitamente, há muitos anos, em escolas como a New Trier High School².

Desde 1923. O. Decroly e R. Buyse preconizaram esse sistema «de classes móveis, tendo em conta as diferenças de aquisição em certas disciplinas»³.

Notemos, no entanto, que os dados experimentais de que dispomos actualmente parecem indicar que o recurso aos grupos homogeneizados não tem realmente interesse a não ser que os

¹ Sistema que não deve confundir-se com o *streaming* que consiste em formar classes homogéneas mediante testes de aptidões gerais ou, pior ainda, em função dos resultados escolares globais. Comete-se deste modo um duplo erro científico. Na grande maioria dos casos, uma mesma pessoa é desigualmente dotada para as diversas disciplinas do programa; além disso, homogeneizar globalmente faz quebrar o rendimento do conjunto. Com este sistema termina-se muitas vezes por uma semiseparação social. Todos estes inconvenientes foram tão bem postos em evidência que o *streaming* foi interdito por lei na Suécia.

² G. DE LANDSHBERE, *L'individualisation de l'enseignement dans une école multilatérale*, in Education, Maio de 1961.

³ O. DECROLY e R. BUYSE, *Les applications américaines de la psychologie à l'organisation humaine et à l'éducation*, Bruxelles, Lamertin, 1923, p. 45.

professores variem a estratégia pedagógica em função dos grupos, e não se contentem apenas com modificar o nível de exigência sem mudar mais coisa nenhuma.

Previnamos imediatamente a objecção financeira. Uma escola flexível não custa necessariamente mais do que outra qualquer. Pelo contrário. Se somarmos o que se poupa em tempo, talento e a diminuição de maus resultados que o sistema permite, poderemos afirmar sem receio de nos enganarmos que é considerável o benefício...

Que aprendizagens escolares possam ser dominadas pelo maior número de alunos, senão por todos, não deve, no entanto, levar-nos a acreditar na desapareição das diferenças entre indivíduos.

Com efeito, as aplicações práticas de métodos geradores de domínio são ainda muito raros e muito recentes para podermos dispor de dados precisos sobre o futuro das populações que deles teriam vindo a beneficiar. Parece razoável admitir que se elevará a média geral das aquisições, mas também que cada indivíduo ao encontrar possibilidade de melhor desenvolver tudo o que existe nele se virá a diferenciar mais nitidamente dos outros do que em tempos idos. Em resumo, ao lado da curva em j , correspondente às aprendizagens fundamentais, a curva de Gauss representará talvez melhor que nunca a distribuição dos talentos individuais.

CAPÍTULO 4

A TEORIA DA EVOLUÇÃO SOMATIVA

A avaliação somativa vai buscar à avaliação formativa um certo carácter diagnóstico, mas é muito mais geral. A avaliação diz respeito, neste caso, seja a uma parte importante dum curso, seja mesmo a todo o curso. Falamos de cursos por preocupação de clareza; mas pode tratar-se também de objectivos gerais tais como a aquisição duma atitude ou duma capacidade (saber desempanar um automóvel).

Todavia, os exemplos que propomos são enganadores no sentido de incidirem sobre vastos conjuntos ao passo que nem sempre assim acontece. *A avaliação somativa* pode incidir apenas sobre a subtracção (ver exemplo de Hively, p. 260).

A distinção fundamental reside menos na grandeza da aprendizagem considerada que no objecto da avaliação. *A avaliação somativa traduz-se por um score.*

Para o educador, o carácter diagnóstico individual da avaliação somativa reveste-se duma grande importância. Todavia, seria lamentável ignorar as possibilidades de comparação entre indivíduos ou grupos seguindo um mesmo programa, ou até entre indivíduos ou grupos seguindo programas diferentes. É talvez mesmo o único modo verdadeiro de obter comparações realmente significativas. Voltaremos a este caso.

Chegados a este ponto, é essencial recordar a diferença entre os *testes normativos* e os *testes centrados sobre os objectivos* (*criterion referenced tests*). Os primeiros têm por objecto situar uma *performance* (uma aprendizagem) em relação às *perfor-*

mances dum conjunto de indivíduos. A distribuição destas *performances* constitui a norma.

Os testes centrados nos objectivos têm também uma norma. Mas essa norma é constituída pelo conjunto (ou um subconjunto) das componentes duma aprendizagem a efectuar. Por exemplo, saber traduzir em inglês 1475 palavras do *Francês fundamental* (1.º grau) poderia constituir um critério, e as notas atribuídas ao aluno seriam o número de palavras conhecidas.

Vê-se imediatamente o parentesco entre a avaliação formativa e a avaliação somativa.

Todavia, os conceitos não se sobrepõem inteiramente.

Poderá ajudar-nos uma *lapalissada*: para que haja *avaliação somativa*, é necessário que haja *somatório de unidades* de aprendizagem, ao passo que um teste centrado sobre objectivos pode muito bem comportar apenas uma dessas unidades.

Lembraremos sobretudo que avaliação formativa e testes centrados em objectivos procedem duma mesma opção educativa: considerar o indivíduo em desenvolvimento e não em competição com os outros.

Para a redacção dos *items*, vimos (p. 122) as principais posições dos subjectivistas e dos objectivistas.

No domínio dos testes somativos, surge um problema semelhante. Merece também alguma atenção em virtude da sua importância na teoria da construção dos programas escolares (*curriculum development*) ou, mais modestamente, nas investigações que visam a controlar a eficácia dos diferentes métodos de ensino.

Parece evidente que, para testar os efeitos dum programa, as perguntas ou tarefas a efectuar devem inscrever-se nesse programa. Interrogar sobre o que não se ensinou parece aqui injustificável¹.

¹ O *transfert* das aprendizagens põe evidentemente um terrível problema de limites. Abstraímos dele momentaneamente.

Ora, a avaliação dos efeitos de dois curricula diferentes, com o auxílio dum mesmo instrumento, levanta quase sempre essa dificuldade. O exemplo mais evidente destes últimos tempos é dado pela comparação dos rendimentos das escolas secundárias selectivas e das escolas únicas («compreensivas»). T. Husen escreve e muito bem: «Os critérios de avaliação dos sistemas escolares elitistas ou unificados dependem de juízos de valor: os sistemas de diferenciação apressada tomam não raro por critério de sucesso um produto final particular (por exemplo os conhecimentos em matemática), ao passo que os sistemas unificados insistem no aspecto afectivo da educação, na abertura de espírito, na disponibilidade intelectual e social. Como demonstrar que uma escolha vale mais que a outra?»

Em semelhante caso, a comparação deveria ter lugar em dois tempos. Uma primeira série de medidas tentaria esgotar comparativamente a parte comum dos dois programas. Uma segunda série incidiria distinta e explicitamente nos objectivos específicos que diferem dum programa ao outro.

Para a primeira série, que diz respeito à parte comum, ou se impõe um mesmo teste ou irão aplicar-se dois instrumentos a respeito dos quais se possa demonstrar que cada um deles é a *soma de items* gerados a partir das mesmas regras, em iguais níveis taxonómicos, e resultando todos de tiragens ao acaso no universo das perguntas. Estamos assim perante a teoria de Bormuth, já encontrada.

CONCLUSÕES
E
RECOMENDAÇÕES

As conclusões parciais foram-se impondo à medida que avançamos neste estudo já longo. No entanto, estamos bem longe de ter esgotado o assunto e, mais ainda, de ter dado soluções definitivas.

Sair das rotinas irreflectidas e ambicionar traduzir as grandes teorias pedagógicas na prática quotidiana da educação, encontra tantos escolhos, tantas condições a preencher que corremos o risco de nos deixarmos invadir pela dúvida e pelo desânimo.

É preferível deixar os leitores insatisfeitos a ocultar-lhes a dificuldade, ou, pior ainda, a convencê-los de que existe uma docimologia perfeita, capaz de levar a uma avaliação exacta, se pudermos dispor dos meios para tal. Naturalmente, e felizmente sem dúvida, a medida rigorosa das características humanas mais delicadas permanecerá sempre impossível. Deus nos livre da máquina de pesar almas!

Pelo facto de a avaliação contínua se inserir funcionalmente no processo de ensino e de aprendizagem de que não pode dissociar-se, nem por isso dispensa a necessidade duma docimástica, isto é duma técnica de exame e de classificação ou melhor ainda duma técnica de observação sistemática do rendimento imediato ou diferido.

Além disso, a exigência de avaliação comparada, que o educador pode lamentar por razões ideais, ultrapassa o domínio escolar: corresponde a um carácter fundamental da nossa civilização.

Talvez um dia seja de outra maneira. Se a nossa cultura continuar a intelectualizar-se, e portanto a crescer em complexidade, não chegamos a ver como poderia renunciar completamente aos exames e aos concursos.

A nossa posição docimológica é resolutamente eclética no seu desejo de conciliar as vantagens seguras de técnicas novas, de natureza estatística sobretudo, com o contributo inegável e fundamental da análise qualitativa.

Para nos limitarmos a um único exemplo desse ecletismo, não podemos conceber que as decisões de passagem que têm lugar no fim de cada ano dos estudos universitários sejam tomadas em função apenas dos resultados obtidos com um teste de escolha múltipla de que resulta uma classificação automática. Não há dúvida de que a introdução deste tipo de provas é altamente desejável para o controlo objectivo e profundo dos conhecimentos, mas não basta. Tem de intervir a avaliação do trabalho do ano e o encontro final do mestre com o aluno deve revelar todas as cambiantes que os dados quantitativos deixaram na sombra.

Assim como a introdução das máquinas na indústria permitiu que a inteligência fosse substituindo a força muscular e a rotina aviltante, do mesmo modo o controlo automático do conhecimento dos factos, dos métodos e das técnicas torna possível um exame final centrado realmente sobre os processos mentais e sobre as mais nobres manifestações da personalidade. Que importância respectiva será dada aos três tipos de avaliação? Neste caso, a decisão pertence à comunidade académica, esclarecida pelas investigações docimológicas.

Parece impor-se um certo número de *recomendações*.

Em virtude mesmo da nossa ignorância e da grande complexidade da avaliação, é urgente, em primeiro lugar, que *uma rede de centros de investigação em educação* recubra o país inteiro e seja posta ao serviço do sistema escolar. As funções desses centros foram aparecendo no decurso da nossa análise: investigações que incidiam sobre os problemas fundamentais e desenvolvimento dos métodos, das técnicas e dos instrumentos necessários. Estes centros não devem ser enxertados em corpos estranhos ao conjunto das nossas instituições educativas, mas trabalhar em estreita colaboração com elas.

Mais particularmente, *cada universidade* deve criar no seu seio *um gabinete de estudos dos problemas de ensino e de exames*. Não se pode conceber que a instituição científica por excelência exclua do rigor da sua análise uma das suas razões de ser: a educação.

A *definição dos objectivos* gerais e especiais, fundamento de todo o edifício do ensino e da avaliação, requer o trabalho em comum dum grande número de investigadores e de praticantes de todas as disciplinas. *Comissões especializadas* irão permitir a cooperação necessária e a comunicação — nos dois sentidos — entre a base e o vértice da hierarquia pedagógica.

A reforma dos exames e dos métodos de avaliação levará sem dúvida à desapareição definitiva do cálculo da percentagem, quase universalmente difundido nas nossas escolas até aos últimos tempos. Quaisquer que sejam as suas falhas, a escala das percentagens oferecia no entanto a vantagem de ser familiar a todos e cada um tinha ao menos a ilusão de a compreender. Actualmente, foram ensaiadas nas nossas escolas as mais diversas escalas. São indispensáveis, no mais breve espaço de tempo, *uma unificação do sistema de classificação* e *uma campanha de informação*, sem as quais se tornará difícil, senão impossível, a comunicação entre a escola e a família, entre as diferentes escolas e as autoridades. Todavia, antes de tomar uma decisão cuja importância é manifesta, deveriam ser ouvidos os peritos da docimologia e, de modo mais geral, da medida em ciências humanas.

Terá de ser empreendido ainda um vasto esforço de informação junto dos docentes, que deverão mudar profundamente os seus hábitos de avaliação. O meio mais eficaz parece consistir na *organização de experiências locais* tendo como objecto a preparação dos exames, a classificação e a moderação. Essas experiências seriam, por exemplo, incorporadas nos trabalhos necessários à introdução do ensino secundário renovado na Bélgica, e na reforma da formação dos professores. Todas as escolas iriam gradualmente participar.

Serão constituídas pouco a pouco *comissões de moderação locais e regionais*.

Uma acção precipitada pode pôr em perigo uma renovação eminentemente louvável. Evitar-se-á particularmente desiludir ou desorientar os professores, tentando colocá-los sob a autoridade de docimologistas improvisados.

Obra humana por excelência, a educação só ganhará em qualidade na medida em que os educadores venham a aceitar profundamente as ideias novas e tentem ultrapassar-se a si mesmos.

ANEXOS

ANEXO I

ESTUDO COMPARADO DUMA PERGUNTA DE EXAME APRESENTADA SEGUNDO O MÉTODO TRADICIONAL E SEGUNDO O MÉTODO COM PERGUNTAS DE ESCOLHA MULTIPLA

*Redacção da pergunta*¹ (método tradicional).

- 1.º a) Indicar a lista das causas da hipoglicémia.
- b) Descrever as lesões anátomo-patológicas resultantes da hipoglicémia.

Para permitir aos correctores um certo grau de uniformidade na apreciação, foi estabelecido o seguinte quadro pelo conjunto dos professores:

- A. Valor igual para as duas partes da pergunta;
- B. Para obter uma nota de 75/100, o candidato deve ter indicado as causas, na lista:
 - 1) Excesso de insulina;
 - 2) Tumor funcional dos ilhéus de Langerhans;
 - 3) Influência endócrina por hipofuncionamento da hipófise e das suprarenais;

¹ MOORE, Robert A., *Methods of Examining Students in Medicine*, in «Journal of Medical Education», Janeiro de 1954, vol. 29, n.º 1.

Tradução extraída do relatório francês sobre a reforma dos exames de medicina, o. c.

- 4) Doença do fígado que impede a formação de reservas do glicogénio (necrose aguda) ou a libertação do glicogénio (doença de Von Gierke).
- 2.º Em que é que se reconhece a diferença entre lesões causadas por uma única crise aguda de hipoglicémia e as que têm origem numa hipoglicémia crónica?
- 3.º Descrever as modificações visíveis ao nível do cérebro.
- C. Para obter uma nota compreendida entre 75 e 90/100, o candidato deve ter indicado:
- ou uma lista mostrando que compreende o mecanismo de acção de cada uma das causas,
 - ou uma lista que compreenda outras causas, mas indicando compreender que os mecanismos de homeostase têm normalmente tendência a corrigir a hipoglicémia em casos como:
- 1) ingesta diminuídos;
 - 2) absorção perturbada;
 - 3) metabolismo aumentado como no hipertiroidismo;
 - 4) infecção grave;
 - 5) *surmenagem* física;
 - 6) glicosúria renal.
- D. Para obter uma nota superior a 90/100, o candidato deve ter indicado:
- uma lista lógica das quatro causas principais e das seis causas menos importantes, além de um tipo ideopático;
 - uma diferenciação da hipoglicémia aguda e crónica mediante uma descrição pormenorizada das lesões cerebrais.

Explicação da pergunta (método das perguntas de escolha múltipla)

Vejam como uma mesma pergunta poderia ser apresentada segundo o novo método.

- 1.º Podemos determinar se o candidato é capaz de reconhecer as quatro causas mais importantes e diferenciá-las das menos importantes na pergunta seguinte:

INSTRUÇÕES — Quatro das cinco frases numeradas são comuns a uma das três perturbações funcionais da lista alfabética (A.B.C.).

Indique a que constitui excepção e a perturbação funcional comum às quatro frases restantes.

- | | |
|---|--|
| 1. Excesso de insulina | A. Hipoglicémia clinicamente detectável |
| 2. Tumor funcional dos ilhéus de Langerhans | B. Hiperglicémia clinicamente detectável |
| 3. Glicosúria renal | C. Glicosúria clinicamente detectável |
| 4. Hipofuncionamento hipofisário | |
| 5. Doença de Von Gierke | |

Se o candidato responder correctamente, mostra saber que 1, 2, 4 e 5 podem produzir uma hipoglicémia clinicamente detectável, o que não é o caso de 3, e que nenhuma combinação dos quatro casos de entre os 5 considerados pode ser associada nem à hiperglicémia nem à glicosúria. Noutros termos, mostra assim que tem conhecimentos positivos e negativos.

- 2.º Se pretendermos verificar o que cada candidato sabe sobre os fenómenos que comandam o equilíbrio do nível de açúcar sanguíneo e de insulina, poderemos fazer a pergunta seguinte:

INSTRUÇÕES. Cada uma das seguintes frases é composta de duas partes: a exposição de um facto e a justificação desse facto.

Indique, para cada uma das frases enumeradas, a letra A, B, C, D ou E, tendo em conta o seguinte:

- A. a exposição do facto e a sua justificação são verdadeiras ou têm uma relação de causa a efeito;

- B. a exposição do facto e a justificação são verdadeiras, mas não há relação de causa a efeito;
 - C. a exposição do facto é verdadeira, mas a justificação é falsa;
 - D. a exposição do facto é falsa, mas a justificação é um facto ou um princípio assente;
 - E. A exposição do facto e a justificação são falsas.
- (A) 1. A taxa do açúcar sanguíneo cai brutalmente depois da hepatectomia, porque o glicogéneo contido no fígado é a principal origem do açúcar sanguíneo.
 - (E) 2. O diagnóstico anátomo-patológico dum adenoma das ilhotas de Langerhans do pâncreas implica que a doença era hiperinsulínica porque todos os adenomas dos ilhéus são funcionais e segregam insulina.
 - (C) 3. Os doentes atingidos de hipertiroidismo têm todas as probabilidades de sofrer duma hipoglicémia, porque existe um hiperinsulinismo associado.
 - (C) 4. Os doentes atingidos da doença de Von Gierke apresentam uma diminuição da taxa do açúcar sanguíneo, porque, nesta doença, o glicogénio não é posto em reserva pelo fígado.
 - (E) 5. Uma hipoglicémia que perdura há muitos meses não é seguida de sequelas, porque as modificações celulares produzidas pela hipoglicémia são reversíveis.

Se o candidato responder correctamente a esta série, mostra que conhece:

- a) que o glicogénio hepático é a fonte principal que permite a manutenção da taxa do açúcar sanguíneo;
- b) que nem todos os tumores dos ilhéus de Langerhans são funcionais;
- c) que os doentes atingidos de hipertiroidismo têm uma hipoglicémia, mas que a causa não é um hiperinsulinismo associado;

- d) que os doentes atingidos pela doença de Von Gierke têm uma hipoglicémia, mas que esta não é devida ao facto de não haver glicogénio no fígado;
- e) que há sequelas após a hipoglicémia crónica e que as modificações celulares resultantes não são reversíveis.

3.º Podemos determinar se o candidato tem algumas noções simples no que diz respeito a modificações ao nível do cérebro, mediante a pergunta seguinte:

INSTRUÇÕES. Cada uma das exposições incompletas (numeradas) é seguida de cinco complementos à escolha. Indique, em cada caso, o complemento que melhor convém.

1. As modificações anatómicas secundárias a uma hipoglicémia crónica são postas em evidência, as mais das vezes:
 - a) no baço
 - b) nos rins
 - c) nas suprarenais
 - d) no cérebro
 - e) na tiróide.
2. De entre as modificações provocadas pela hipoglicémia crónica ao nível do cérebro, as mais importantes são:
 - a) nos neurónios
 - b) nos astrocitos
 - c) nas células do epêndima
 - d) nas células oligodendrógicas
 - e) nas células micrógicas.
3. Dentre as alterações das células nervosas provocadas pela hipoglicémia aguda, a mais evidente é:
 - a) o deslocamento do núcleo
 - b) a desapareção da parede celular
 - c) a alteração da substância de Nissle
 - d) o rebentamento do núcleo da célula
 - e) a fusão das mitocôndrias.

4. Dentre as modificações provocadas ao nível do cérebro pela hipoglicémia crónica, a mais significativa é:
- a) a hidrocefalia interna
 - b) o espessamento fibroso da aracnoide
 - c) a destruição dos neurónios
 - d) a proliferação dos astrocitos
 - e) a proliferação das células do epêndima.

Se o candidato responder correctamente a esta série, mostra que sabe que as alterações principais da hipoglicémia atingem os neurónios do cérebro, que a hipoglicémia aguda provoca uma alteração da substância de Nissle dos neurónios e que a hipoglicémia crónica provoca a destruição das células nervosas.

- 4.º Pode-se determinar se o candidato compreende o mecanismo das perturbações, mediante a pergunta seguinte:

INSTRUÇÕES. Na lista alfabética são indicados cinco mecanismos diferentes que podem conduzir à hipoglicémia. Pôr a letra apropriada depois de cada uma das exposições numeradas, associando à perturbação o mecanismo responsável.

- A. Aumento da actividade metabólica.
- B. Hiperinsulinismo.
- C. Depósito em reserva dum glicogénio anormal no fígado.
- D. Ausência de depósito em reserva de glicogénio no fígado.
- E. Hipofuncionamento da hipófise ou das suprarrenais.

- (B) 1. Adenomas dos ilhéus de Langerhans.
- (A) 2. Exercício físico violento.
- (A) 3. Hipertiroidismo.
- (E) 4. Doença de Simmonds.
- (C) 5. Doença de Von Gierke
- (D) 6. Hepatite epidémica.
- (C) 7. Doença de Addison.

Se o candidato responder correctamente a esta série, mostra que compreende as bases do metabolismo dos glúcidos e conhece os factores que influenciam esse metabolismo.

Recapitulação

Deste modo, mediante estas dezassete «perguntas objectivas», puzemos em evidência todos os conhecimentos requeridos para poder dar uma nota superior a 90/100, ou seja:

1. Lista das quatro causas principais.
2. Reconhecer a diferença entre os efeitos da hipoglicémia crónica e aguda.
3. Descrever as alterações ao nível do cérebro.
4. Compreensão dos mecanismos.
5. Lista das causas menores.
6. Diferenciar as lesões da hipoglicémia aguda e da crónica.

ANEXO II

EXEMPLO DE PERGUNTAS PARA UMA COMPOSIÇÃO DE LÍNGUA MATERNA ¹

PROVA I. Duas perguntas — 1½ h.

Perguntas

1. Escolha um dos temas seguintes. Dedique-lhe cerca de 1 h.
 - a. Uma noite de bruma.
 - b. Um comerciante compra e revende, no mesmo dia, um artigo de saldo. Descreva as duas cenas.
 - c. Qual seria a sua política se dirigisse os programas de rádio ou de televisão?
 - d. Uma grande multidão a dispersar-se. Descreva a cena.
 - e. O prazer da fotografia ou do desenho ou da dança ou do ciclismo.
 - f. Pensa que rapazes e raparigas têm as mesmas probabilidades de obter um emprego?
 - g. Quais são as suas reacções em face do progresso e das realizações da exploração espacial?
2. Escolha um dos temas seguintes. Consagre-lhe cerca de ½ h.
 - a. Descreva um treino, destinado a melhorar as suas *performances* num desporto à sua escolha.

¹ Inglaterra — General Certificate of Education, 1967. Fim do secundário — Nível comum.

- b. Depois de ter visitado uma empresa, faça aos seus companheiros um relatório sobre as condições de trabalho e as perspectivas de futuro que ela oferece.
- c. Descreva claramente *um* dos aparelhos seguintes e explique como ele funciona: um «walkie talkie», um secador de cabelo, um aspirador, uma máquina bate-deira (mixer).
- d. Na cidade onde vive pretende-se criar um centro comercial em que seja interdita a circulação a veículos. Escreva uma carta a um jornal local a expor os seus pontos de vista sobre esse projecto.

PROVA II. Quatro perguntas — $1\frac{3}{4}$ h.

1. Resuma a passagem seguinte em boa prosa contínua e com o máximo de 110 palavras. Finalmente indique quantas palavras utilizou. A passagem consta de 314 palavras.
Segue-se um texto sobre a confiança exagerada na ciência e na tecnologia.
2. Leia a seguinte passagem; depois responda às perguntas. (O texto descreve dois grandes tipos de excursionistas — os que seguem um guia e os que partem à aventura — e sublinha o interesse das excursões geológicas).
 - a. Explique com palavras suas a diferença entre os dois tipos de excursionistas.
 - b. Qual o conselho que o autor dá a propósito das grutas? Formule-o com as suas próprias palavras.
 - c. Explique brevemente porque razão o autor julga que a geologia é: um *hobby* divertido — um *hobby* instrutivo.
 - d. O autor diz que utiliza a expressão «cavar um fosso» em sentido literal. Explique porque é que o sentido é literal aqui.
 - e. Explique as expressões seguintes: precauções prescritas; estar perfeitamente consciente da natureza dos seus actos.

- f. Escolha quatro das seguintes palavras. Substitua-as por sinónimos ou perífrases que poderiam ser utilizadas no texto sem lhe mudar o sentido (...).
3. Responda à sua escolha a uma das seguintes perguntas:
 - a. Escolha três palavras das seguintes. Construa frases (6 ao todo) mostrando que essas palavras podem ser empregadas em dois sentidos diferentes (...).
 - b. Defina numa frase três das palavras seguintes: monópólio, interlúdio, prefácio, microscópio, antídoto.
 4. Responda à sua escolha a uma das seguintes perguntas:
 - a. Explique clara mas brevemente a diferença de sentido entre cada par de frases (incide em *could-should*; *can-may*; *wil-shall*; *might-must*).
 - b. Escreva outra vez correctamente a passagem seguinte, respeitando todas as ideias. Pode mudar a expressão, a ordem das palavras e das ideias, a ortografia e a pontuação.
 (Segue-se um texto defeituoso duma centena de palavras).
 «Por exemplo se se projectasse fazer passar uma estrada através duma cidade, mas houvesse um edifício histórico no caminho, nesse caso deveriam alterar-se as planos, implicando despesas consideráveis para contornar o edifício, originando uma curva na estrada e tornando-a assim perigosa para os automóveis. O desvio dessa estrada não só originaria despesas como ainda aumentaria particularmente o preço dos diversos elementos se o desvio for grande, porque se a distância aumentada for de dez quilómetros e um camião transportando certos artigos percorrer a rua, o camião viria a consumir mais gasolina, gastaria mais tempo para chegar e perderia dez quilómetros de vida».

Indicações para a correcção

PROVA I— Máximo 50 pontos.

O esquema de classificação que figura aqui constitui apenas um guia preliminar. Poderão ser feitas adições e correcções na altura da reunião dos examinadores que será convocada depois duma primeira leitura de trabalhos.

Pergunta 1 (máximo 35 pontos).

É de esperar um mínimo de 400 palavras, mas as composições não devem ser classificadas predominantemente em função da extensão. Ter em conta o assunto escolhido e a maneira como é tratado. Mesmo que seja breve, uma composição em que a argumentação é serrada e a expressão é boa deve obter mais pontos do que uma longa narração informe.

O examinador deveria ter uma ideia clara do que é uma composição que merece precisamente a nota de passagem (16 pontos). Uma semelhante composição deve contar ideias razoáveis, mas não muito originais. A expressão deve ser clara, mas sem distinção particular. No trabalho apenas serão de admitir alguns erros mecânicos. Os candidatos que ultrapassaram o nível geral devem ser recompensados e os que ficaram abaixo desse nível devem ser penalizados.

Apresentamos no documento anexo notas pormenorizadas sobre as qualidades a obter nas composições. É de contar no entanto que os examinadores classificarão tendo em conta a sua impressão geral, não adoptando portanto uma proporção fixa de pontos para os diferentes aspectos. Se o assunto se presta a controvérsia, as ideias e a sua articulação podem ser mais importantes do que numa composição descritiva em que o vocabulário poderia tomar maior relevo. Confiamos na capacidade de julgar dos examinadores.

Reservando 35 pontos para esta primeira pergunta, pretendeu-se assinalar a sua importância dominante no conjunto dos exames. Muitos candidatos são medíocres. Em todo o caso,

se o examinador não dispersar muito as suas notas, esta primeira pergunta não terá um peso suficiente no conjunto dos resultados.

Grupo A (29-35 pontos).

A composição tem qualidade excepcional.

Grupo B (22-28 pontos).

A qualidade do trabalho é acima da média.

Grupo C (14-21 pontos).

O trabalho é de qualidade média.

Grupo D (7-13 pontos).

O trabalho é de um nível inferior ao considerado como satisfatório.

Grupo E (0-6 pontos).

O candidato é incapaz de apresentar as ideias com coerência.

Esta prova é um exame de língua materna. Na prova 1, examinamos a capacidade de o candidato exprimir as suas opiniões, as suas experiências, as suas impressões, os seus sentimentos e os seus interesses. Não se trata nem dum teste de conhecimentos gerais nem duma avaliação das aptidões com clareza uma cena em inglês, merece uma nota favorável se os materiais utilizados forem pertinentes, mesmo que o examinador julgue que a dita moça aborda o assunto por forma demasiado sentimental. Além disso, se o examinador tem a impressão que, por traz duma massa de prosa incoerente, há emoções profundas e atitudes morais elevadas, não tem que se preocupar com este último aspecto. Só conta o que foi escrito. Para o bom êxito da prova são essenciais a clareza de expressão e a precisão do estilo.

Pergunta 1 (máximo 35 pontos)

- (a) Apresentam-se como aceitáveis uma narração ou uma descrição. Numa narração, deverá desempenhar um papel essencial a noite de bruma.

- (b) As duas cenas, de um modo geral, devem ser equilibradas. Recompensar a vivacidade de narração, de diálogo de descrição e o contraste.
- (c) É de esperar uma definição clara da política segundo a qual vários pontos deverão ser desenvolvidos. O candidato poderá ter escolhido a rádio ou a televisão; não as duas em conjunto.
- (d) O tema do trabalho é a dispersão; autorizar no entanto uma breve introdução.
- (e) Uma resposta coerente, clara, de extensão moderada será de preferir a longos devaneios e repetições.
- (f) Este tema não é fácil: recompensar generosamente a boa ordenação dos argumentos e os exemplos bem escolhidos.
- (g) Não pode aceitar-se um tratamento puramente narrativo, mas pode haver necessidade de alguns exemplos de progresso realizados para explicar as reacções.

Pergunta 2 (máximo de 15 pontos).

- (a) Pretende-se uma exposição clara e lógica.
- (b) O relatório deve tratar os três aspectos do problema. Insistir-se-á sobretudo no tipo de informação aduzida. Não se irá exigir, portanto, uma forma especial de relatório.
- (c) Exigir-se-á a um tempo uma descrição do objecto e uma explicação do seu funcionamento. Não serão de considerar com severidade os erros materiais. Encarar simplesmente a clareza de expressão.
- (d) Poderemos ter em conta diversos aspectos. No entanto, constituirá sempre desvalorização qualquer ideia inadequada.

Descontar:

1 ponto para uma má redacção da mensagem.

1 ponto para uma falta de coerência entre a vedeta e as saudações.

1 ponto, se o aluno assinou «Sr. João Smith» ou «Sr.^a D. Joana Smith».

$\frac{1}{2}$ ponto para outros erros de disposição ou de pontuação ou de ortografia das palavras essenciais.

Todos estes temas fornecem um material suficiente para uma meia hora de trabalho. É de prever um mínimo de 200 palavras. A escolha é rica.

PROVA II — Máximo de 50 pontos.

Pergunta I (máximo de 16 pontos).

Os pontos atribuídos a esta pergunta são muito mais baixos habitualmente do que os atribuídos às outras. Pede-se aos examinadores para não considerarem 11 como o máximo que se pode atribuir ao resumo.

A. — Atribuir um máximo de 2 pontos a cada um dos seguintes aspectos. Para obter dois pontos, o candidato deve ter compreendido a ideia claramente, exprimindo-a com correcção. Distribuir-se-ão as notas em $1\frac{1}{2}$, 1 ou $\frac{1}{2}$ pontos. No fim do trabalho, conte de novo o conjunto em função da fluidez e da coerência do resumo completo. Se uma passagem for incoerente, suprima quando muito $\frac{1}{4}$ dos pontos atribuídos. Se se apresentar sobretudo pouca cuidada, suprima $\frac{1}{8}$. Essa correcção deveria figurar na composição, escrevendo por exemplo $10 - 1 = 9$. Toda a subtracção de pontos motivada por um texto mais longo que o limite fixado ou uma afirmação não apenas inexacta mas absurda deve aparecer separadamente. Deverá envolver-se com um círculo à margem o total final.

- 1.º O homem da rua aceita hoje as descobertas científicas.
- 2.º sem duvidar da sua origem, da sua validade, ou dos seus efeitos.
- 3.º e 4.º A exigência de novos progressos destinados a elevar o nível de vida não afrouxa nunca.
- 5.º Os homens têm confiança no homem de ciência e no seu trabalho,
- 6.º e julgam que não é possível deter o progresso.

- 7.º Embora reconheçam que os homens de ciência nem sempre estão de acordo no que toca à tranquilidade que as novas descobertas trazem,
- 8.º o público está convencido de que os homens de ciência acabarão por encontrar a unanimidade ou pelo menos uma grande concordância.

B. — O limite de 120 pontos dá uma margem generosa. Suprima um ponto por cada conjunto de 5 palavras que ultrapassem o máximo. Não inclua na contagem palavras introdutórias como «Nesta passagem, o autor explica que...»; Torne a contar as palavras. Não aceite pura e simplesmente o número inscrito pelo candidato.

Pergunta 2 (máximo de 20 pontos).

Ao classificar esta pergunta, dê o máximo em cada secção ao candidato que exponha claramente o assunto.

- (a) 1. Os que gostam que tudo esteja previsto e tudo no seu lugar (1)
2. e os que preferem dispor apenas dum plano geral que permita seguir a inspiração do momento (espírito de exploração) (2)
- (b) Visitar grutas (1)
- A não ser que vos sirva de guia alguém que conhece muito bem o terreno. (1)
- Respeite todas as regras de segurança. (2)
- (c) 1. É dispensável equipamento especial (2)
2. O geólogo faz a cada passo novas descobertas em pequena escala. ($1\frac{1}{2}$) ($1\frac{1}{2}$)
- (d) Quando o geólogo parte uma pedra, executa de facto um acto único, porque ninguém depois dele poderá partir de novo essa mesma pedra (2)

- (e) 1. Medidas de segurança impostas ou recomendadas (1)
(1)
2. Para compreender exactamente as consequências do que se faz (1)
(1)

(f) Subterrâneo:

que está debaixo do solo, debaixo da superfície da terra; (1)
debaixo da terra. (1)

Espeleólogos:

peçoas que estudam as grutas cientificamente; (1)
exploradores de grutas. (1)

Conscienciosamente:

de boa fé, sem se deixar distrair, com muita aplicação (1)
honestamente (1/2)
fielmente (1/2)

Etc.

Para as sub-perguntas (a), (b) e (c) não atribua nenhum ponto aos alunos que se limitarem a recopiar uma parte do texto.

Pergunta 3 (máximo 6 pontos).

- (a) Atribua um ponto a toda a frase construída correctamente. Não atribua nenhum ponto em caso de construção incorrecta.
- (b) Dê um ponto a cada definição exacta expressa numa frase correcta (máximo 3 pontos).
Dê um ponto a cada frase em que o sentido da palavra que começa pelo mesmo prefixo ressalta claramente (máximo 3 pontos).

$\frac{1}{2}$ ponto para uma definição **expressa por uma frase** incorrecta.

$\frac{1}{2}$ ponto para uma definição expressa por uma frase incompleta.

Zero ponto se a palavra que começa pelo **mesmo prefixo** não for apresentada numa frase.

Um *monopólio* é uma propriedade exclusiva mantida por uma firma (ou)

Um *monopólio* é o nome dado a uma firma que possui direitos comerciais exclusivos.

Um *interlúdio* é um intervalo que ocorre no decurso da representação duma peça (ou um intervalo durante o desenrolar dum acontecimento).

Pergunta 2 (máximo 8 pontos).

(a) Dê um ponto pela frase

Verifique a presença das seguintes ideias...

Não atribua nenhum ponto se não se vir claramente a que frase se refere o candidato.

(b) Os candidatos irão responder de modos diversos.

Deduz a 1 ponto por falta de ortografia, 1 ponto por uma expressão incorrecta, $\frac{1}{2}$ ponto por omissão ou emprego errado duma vírgula *essencial*, 1 ponto por cada ideia omitida.

Reveja a nota total em função da impressão geral da passagem.

Para as perguntas 1 e 2, se o candidato obtiver metade dos pontos, arredonde para a unidade superior (ex.: $6\frac{1}{2} = 7$)

Se as duas perguntas derem uma nota que compreenda $\frac{1}{2}$ ponto, arredonde um por excesso e o outro por defeito.

As mesmas observações para as perguntas 3 e 4.

ANEXO III

EXEMPLO DE ENSINO SEMI-INDIVIDUALIZADO¹

A New Trier Township High School, Winnetka

A *New Trier Township High School* é uma grande escola do grau secundário superior em que o ensino é semi-individualizado e em que o estudante pode corrigir a sua orientação até ao fim da adolescência.

Recebe os alunos a partir dos 14 anos² e goza duma grande reputação tanto nos Estados Unidos como no estrangeiro.

3740 estudantes seguem regularmente estes cursos³. O corpo docente é composto de 255 membros dos quais:

40	professores de Inglês
31	» de Matemáticas
30	» de Educação Física
29	» de Línguas Estrangeiras
27	» de «Estudos Sociais» (História, Geografia, Educação Cívica, Sociologia e Economia).

¹ Extracto dum artigo que publicámos em *Education*, Maio de 1961.

² A organização do ensino nos estados Unidos não é uniforme, como se sabe. A N. T. High School depende do Sistema «NK 8-4»: um ano de «Nursery school», um ano de jardim de infância, 8 anos de primária e 4 anos de secundária. Os dois outros sistemas mais frequentes são: NK 6-3-3 e NK 6-6. Também aparecem NK 7-5, NK 6-2-4 e NK 6-4-4.

³ Estes números referem-se ao ano escolar 1959-1960. São extraídos de *Information for College Admission Officers*, N. T. Township High School, Nov. de 1959, onde foram directamente recolhidos.

17	»	de Ciências
9	»	de Música
7	»	de Cursos Técnicos
7	»	de Comércio
6	»	de Arte Dramática
6	»	de Pintura e Desenho
6	»	de Estudos sobre o Automóvel
4	»	de Economia Doméstica
2	»	de Higiene
1	»	de Ensino Especial (retardados mentais educáveis).

Conta-se um docente para 15-16 estudantes e as turmas reúnem geralmente cerca de 25 alunos.

O trabalho dum professor que não assume responsabilidades especiais (tais como o presidente duma secção, por exemplo) consiste em 24 períodos de 40 minutos por semana: 4 cursos de 5 períodos e 4 períodos de *counseling*.

O nível intelectual dos estudantes, julgados na base dos testes clássicos, é elevado. Cerca de 80% dos que terminam os estudos em New Trier obtêm resultados superiores à média nacional, nos testes de aptidão e de conhecimentos (SCAT, STEP, *National Merit*). 92% dos diplomados continuam estudos superiores.

A individualização dos programas

O princípio fundamental da acção pedagógica da New Trier High School é definido na primeira frase do seu programa de curso: «Ninguém pode acreditar na dignidade dos homens sem sentir um prazer profundo perante o espectáculo da sua variabilidade infinita»¹.

Permanecendo assim impregnados do espírito do Plano de Winnetka, os estudos são organizados de modo a fornecer a

¹ NEW TRIER TOWNSHIP HIGH SCHOOL, *Curriculum Guide*, Dezembro de 1959, pág. 1.

a possibilidade de se desenvolver ao seu próprio ritmo, segundo as suas capacidades e tendências. Isto não significa de maneira nenhuma que sejam permitidas tanto a fantasia estéril como as soluções de facilidade.

De acordo com o sistema vigente em todo o país, não poderá o estudante obter um diploma final, a não ser que, no decurso dos seus estudos, tenha conseguido um número total de pontos (ou *credits*) fixado pelo conselho de administração da escola que por sua vez tem em conta normas gerais. Estes pontos exprimem numa unidade convencional a importância quantitativa e qualitativa dos diferentes cursos, para um semestre (5 *credits* ou unidades correspondem, por exemplo, a um semestre de curso «menor», à razão de cinco períodos por semana). Os pontos só serão adquiridos se o estudante atingir uma nota superior a um limite mínimo fixado.

Para obter o diploma de estudos da New Trier High School, deverão ter obtido 350 pontos em 4 anos.

O jogo dos «créditos» permite a um tempo maleabilidade e segurança, porque permite delimitar exactamente o campo de liberdade do aluno.

Em princípio, o estudante estabelece o seu programa de estudos como se compõe uma ementa de refeição, escolhendo na carta o que mais atrai e melhor convém.

Eis, por exemplo, a lista dos cursos oferecidos no primeiro ano, mencionando «créditos» que merecem, quando seguidos com êxito durante um semestre ¹.

Cursos «maiores»

(Dão direito a 10 «unidades» por semestre; salvo indicação contrária; deverão ser seguidos durante dois semestres consecutivos).

¹ Cf. NEW TRIER TOWNSHIP HIGH SCHOOL, Registration Bulletin for Freshmen, 1959-1960. Planning a Course of Study, p. 2.

Inglês	Matemáticas Gerais
Álgebra	Pintura-Desenho (2 períodos por dia)
Alimentação (1 semestre)	Fotografia
Civismo	Rádio amador
Comércio	Trabalho em madeira
Desenho industrial	Trabalhos manuais
Electricidade (1 semestre)	Vestuário (1 semestre)
História Universal	
Línguas estrangeiras	

Cursos «menores»

(Número de «créditos» entre parêntesis)

Arte dramática	(6)	Desenho Industrial (1 período por dia)	(5)
Canto coral	(3)	Dicção	(6)
Dactilografia	(5)	Harmonia	(5)
Curso de Harmonia	(6)	Orquestra Sinfónica	(5)
		Trabalhos Manuais (1 período por dia)	(5)

Todavia, são muitas as restrições que influenciam a escolha do aluno. Em primeiro lugar, é obrigatório um certo número de cursos indispensáveis à cultura base de todos os membros da nação.

Cursos	Duração obrigatória	Pontos
Língua Materna	4 ano	80
Matemática	2 anos	40
«Social Studies»	2 anos ¹	40
Ciências	1 ano	20
Educação física	4 anos	16
Automóvel: teoria e condução	1 semestre	3
		<hr/> Total 199

¹ Dos quais um ano obrigatoriamente consagrado à história dos Estados Unidos.

Cerca de 2/3 dos créditos exigidos são portanto fornecidos por cursos impostos. Todavia, como veremos, é possível satisfazer de forma livre a essas exigências. E assim conseguimos encontrar uma dúzia de possibilidades diferentes que permitem vencer com proveito o ano de ciências em questão.

Um segundo e importante factor guia o estudante na elaboração do seu programa: a profissão ou os estudos superiores a que ele aspira. Em particular, cada universidade determina as condições de admissão e especifica designadamente quantas unidades o estudante deve ter obtido no ensino secundário, para determinadas disciplinas.

Enfim, pais, professores, conselheiros pedagógicos e orientadores esforçam-se por guiar o aluno de acordo com os seus verdadeiros interesses, procurando que tire o maior proveito possível das suas potencialidades.

Como a semana escolar conta 5 dias de 8 períodos de 40 minutos — não compreendidas a reunião quotidiana com o conselheiro pedagógico (20') e o tempo da merenda sempre comida na escola (25') — sugere-se ao aluno que forme um programa medianamente carregado que lhe deixe tempo disponível para o estudo pessoal. Damos quatro exemplos tipo para o primeiro ano ¹.

I		II		III		IV	
Cursos	per. sem.	Cursos	per. sem.	Cursos	per. sem.	Cursos	per. sem.
Inglês	5	Inglês	5	Inglês	5	Inglês	5
Álgebra	5	Álgebra	5	Álgebra	5	Álgebra	5
Latim	5	Ciências	5	História	5	História	5
Pint./des	10	Comércio	5	Aliment.	7	Ciências	5
Ginást.(meninas)	4	Ginást.(rapazes)	5	Canto	3	Língua est.	5
	—	Música inst.	5	Ginást.(meninas)	4	Ginást.(rapazes)	5
	29		—		—		—
Estudo pessoal	11		30		29		30
	—		10		11		10
	40		—		—		—
			40		40		40

¹ NEW TRIER TOWNSHIP HIGH SCHOOL, *Courses for Freshmen*, Abril de 1959, p. 4.

É fácil notar como estes planos de trabalho estão em contraste com a dispersão do esforço tão frequente entre nós. Praticamente, todos os cursos académicos podem concentrar-se na parte da manhã.

Cerca de um quarto do tempo passado na escola é reservado aos estudos pessoais. As horas assim livres serão consagradas a trabalhos de investigação na biblioteca escolar que, nos Estudos Unidos, desempenha um papel incomparavelmente mais importante que nos nossos países.

A individualização do ensino

O aluno não só escolhe a secção que lhe convém mais, mas ainda o ensino de cada uma delas será adaptado às suas possibilidades.

Com efeito, todas as secções importantes podem ser estudadas a cinco níveis de aptidões diferentes: inferior, médio, fraco, normal, acelerado, avançado. Deste modo, o esforço exigido ao estudante, quer possua uma inteligência superior ou seja pouco dotado, quer seja forte num domínio e atrasado noutra, está sempre em proporção com as suas possibilidades.

Como o aluno raramente segue todos os cursos a um mesmo nível, parece mínimo o perigo duma segregação geral, segundo as aptidões. Aliás, a escola está sempre atenta a esse problema, reagrupando sistematicamente todos os estudantes por ocasião de certas actividades. Envidam-se também todos os esforços para inculcar uma verdadeira tolerância em face do maior ou menor talento dos companheiros de estudo e para os encorajar a todos a irem mais além. «No desporto, nem todos são capazes de serem campeões. O mesmo acontece nos estudos. Mas espera-se de cada um o melhor de si mesmo...»¹

O sistema de ensino de diferentes níveis é praticado em New Trier, desde o fim da primeira guerra mundial, com o

¹ NEW TRIER HIGH SCHOOL, *Guide Book to New Trier, Winnetka*, 1959, p. 39.

maior agrado de todos. Os estudantes repartem-se em média segundo as percentagens:

— superiores	15-20%
— normais	40-44%
— médios fracos	36-40%
— limitados	5- 8%
— avançados (seniores)	±10%

Quando os alunos entram na New Trier High School, os grupos determinam-se provisoriamente na base dos resultados escolares anteriores e doutras informações reunidas segundo um sistema que estudaremos depois. Em seguida, os resultados obtidos na própria escola irão corrigir e guiar as colocações que nunca são definitivas, seja qual for a disciplina.

Semelhante maleabilidade conduz sem dúvida a uma organização complexa, mas não devemos exagerar a dificuldade.

No seguinte quadro, mostramos todas as possibilidades oferecidas durante o segundo semestre do ano escolar de 1959-1960. Algumas notas marginais precisam o espírito desses estudos¹.

Cada curso é designado por um número de três algarismos.

- a) O algarismo das centenas indica em que ano é dado. Teremos assim 1, 2, 3 e 4.
- b) O algarismo das dezenas indica o semestre.
 - 1 = 1.º semestre
 - 2 = 2.º semestre
 - 0 = pode seguir-se ao 1.º ou ao 2.º semestre
 - 3 = curso de férias no verão.
- c) o algarismo das unidades indica o nível de aptidão
 - 1 = nível inferior
 - 2 = médio fraco
 - 3 = normal
 - 4 = ensino acelerado
 - 5 = avançado
 - 6 = seminários
 - 9 = todos os níveis reunidos.

¹ Foram utilizados os dois documentos de base seguintes: NEW TRIER TOWNSHIP HIGH SCHOOL, *Curriculum Guide*, 1959 e *Program of Classes*, 1959-1960.

Aspectos e níveis de curso	Observações
<i>Língua materna</i>	
125	— Para estudantes superiormente inteligentes, ótimos em língua materna (integra a língua materna, a biologia e a história).
124	— Curso enriquecido, fundado nas obras literárias de nível universal.
123	— Curso médio.
122	— Curso médio inferior. Destina-se especialmente aos alunos que não adquiriram ainda um método de trabalho racional.
122 R	— R = «remedial». Destinado aos alunos identificados pelo serviço de <i>testing</i> como apresentando uma deficiência marcada em leitura (compreensão e rapidez) e em ortografia.
121	— Acolhe os alunos que apresentam uma fraqueza geral em todos os aspectos da língua materna.
224 - 223 - 222 - 221	
324 - 323 - 322 - 321	
424 - 424 (+) - 423 - 422 - 421	As diferenças de nível no 2.º, 3.º e 4.º anos são paralelas, <i>mutatis mutandis</i> , às que acima indicamos para o 1.º ano. 323 = jornalismo 424 (+) = «Great Books».
<i>Matemática</i>	
Álgebra A (acelerado) 124	A diferença entre os níveis 2, 3 e 4 reside menos na aceleração que no método de ensino e no aprofundamento.
» — 124	No nível 2, as explicações são pormenorizadas; muitas aplicações; nada de incursões nos domínios vizinhos.
» E (experimental) 124	No nível 3, o estudante deve trabalhar mais por si mesmo; teoria mais rigorosa; enriquecimento.
» — 123	No nível 4, os conceitos são tratados rapidamente; estudo mais aprofundado; numerosas incursões nos domínios vizinhos.
» E 123	Emprega-se o mesmo manual para os 3 níveis e todos os alunos estudam o mesmo capítulo ao mesmo tempo, o que permite, a todo o momento, a passagem dum nível ao outro.
» — 122	
» E 122	
<i>Matemática de base</i> 121	
Matemáticas aprofundadas	
(20 unidades) 225	
Geometria 224	
» 223	
» 222	
Matemática 222	
<i>Matemática de base</i> 221	
Matemáticas aprofundadas	
(20 unidades) 315	
Matemática 324	
Álgebra 323	
» 322	

<i>Aspectos e níveis de curso</i>	<i>Observações</i>
Geometria 303	São obrigatórios 4 semestres de matemática; todavia, dois destes podem ser consagrados à contabilidade ou à aritmética comercial.
Matemáticas aprofundadas 425	
Matemáticas 424	
Álgebra (nível universitário) 403	
Trigonometria 403	
Uso da régua de cálculo 323 (2 unidades; com a álgebra 322)	
«Social studios»	São obrigatórios 4 semestres de «Social Studies». Destes devem ser consagrados 2 à história dos Estados Unidos.
Civismo 123, 122, 121	
História Universal 125, 124, 123	
História: tempos modernos 224	
História: Antiguidade 204	
História: Idade Média 202, 203, 204	
História: Universal 221	
Geografia 222 - 223	
História: tempos modernos 223	
História: Grã-Bretanha 303	
História U.S.A. 324 - 323 - 322 - 321	
História: U.S.A. (aprofundada) 425	
História: contemporânea (Europa) 425	
História: Grã-Bretanha (aprof.) 404	
História: América Latina 403	
História: Extremo Oriente 403	
História: Universal, s. xx 402	
História: Estados-Unidos, s. xx 402	
Civismo 403	
Sociologia 403	
Ciências Económicas 403	
Ciências	Os alunos que desejarem continuar estudos nas Faculdades de Ciências das Universidades serão convidados a seguir 6 semestres de ciências e 8 de Matemáticas.
Biologia 123 - 124 - 125	
Radioamador 123	
Biologia 224 - 223 - 222	
Química 225	
Biologia 323 - 322 - 321	
Electrónica 323	
Física 325	
Química 324-323	
Química (seminário) 426	
Física 425 - 424 - 423	
Ciências 409	
Línguas	
Latim 124 - 123 - 122	
224 - 223 - 222	
324 - 323	
424 - 423	
Alemão 124 - 123	
224 223 223C	
324 - 323	
426	

<i>Aspectos e níveis de cursos</i>	<i>Observações</i>
Russo	124 224 324
Francês	125 124 123 224C 224 223C 223 324C 324 323C 323 324=
Espanhol	125 124 123 122 224 224C 223 223C 222 324 324 = 323 322 424 423
	C indica que se trata da continuação dum curso precedente. = curso de conversação.
<hr/>	
<i>Artes</i>	
Desenho-pintura	129
Trabalho manual	121
Desenho-pintura	229C 229
Trabalho manual	221
Desenho-pintura	329
Cerâmica	329
Joalheria	329
História da Arte	329
Cerâmica	429
Joalheria	429
Pintura-desenho	429
	Todos estes cursos podem ser seguidos à razão de 2 períodos por dia (= curso «maior»: 10 unidades) ou de 1 período por dia (= curso «menor»: 5 unidades ou «credits»).
	Todavia a História da Arte 329 é sempre um curso «maior».
<hr/>	
<i>Comércio</i>	
Comércio geral	122 123
Comércio geral	222
Contabilidade	223
O ponto de vista do consumidor	323
Venda	302
Publicidade	303
Organização Comercial	303
Estenografia	323
Dactilografia	303
Prática de escritório	322
Secretariado	423
Estenografia	423
Direito Comercial	423 403
Prática Comercial	422
	<i>Objectivos deste curso:</i>
	1.º Preparação para os estudos superiores de ciências comerciais;
	2.º Formação geral;
	3.º Preparação para os estudantes que querem trabalhar em <i>part-time</i> no comércio durante os seus estudos universitários;
	4.º Preparação profissional para os alunos que não farão estudos superiores;
	5.º Possibilidade, para os estudantes que procuram a sua vocação, de ver se a carreira comercial os interessaria.
	O 422 = uma reunião de 30' todas as manhãs na escola. O resto consiste em estágios práticos (20 h./semanais).
<hr/>	
<i>Automóvel</i>	
Teoria e prática (condução)	209
	Uma lei de 1955 torna este curso obrigatório em todas as High Schools de Illinois (3 unidades).

<i>Aspectos e níveis de cursos</i>		<i>Observações</i>
<i>Economia Doméstica</i>		Destinado particularmente às raparigas; no entanto: — os rapazes que se interessem pela restauração e pela hotelaria podem seguir os cursos de alimentação; — os rapazes que se interessem pela Arquitectura podem seguir os cursos de decoração interior.
Alimentação	109	
<u>Vestuário</u>	109	
Alimentação	209	
<u>Vestuário</u>	209	
Decoração interior	329	
<i>Cursos técnicos</i>		— O objectivo não é dar uma formação profissional, mas tornar o estudante um «consumidor inteligente dos produtos da indústria» (p. 64). — Cada um destes cursos dá 5 ou 6 unidades conforme o tempo que se lhes consagra. — As universidades reconhecem estas unidades; além disso, pretendem que os estudantes adquiram uma formação técnica.
Desenho industrial	123 122	
Madeira	123 121	
<u>Electricidade</u>	103	
Desenho (Arquitectura)	223	
Madeira	222	
<u>Metalurgia geral</u>	223	
Desenho (mecânico)	323	
Metalurgia geral	323	
Noções de Arquitectura	323	
Leitura de plantas	302	
<i>Música</i>		Trata-se tanto duma instituição para amador como de estudos aprofundados que podem preparar para as carreiras musicais.
Canto	129	
Orquestra	129	
Curso de Armonia	129	
<u>Apreciação musical</u>	129	
Canto	229	
Pequeno grupo coral	229	
<u>Apreciação musical</u>	229	
<u>Composição</u>	329	
Madrigais (grupo coral)	429	
<u>Composição</u>	429	
Coral (selecção de 32 estud.)	029	
Ópera	929	
Grande coral	029	
Harmonia (mais novos - grupo de finalistas - grupo de honra)	029	
Orquestra (seniores)	029	
Conjunto instrumental	029	
Piano	029	
<i>Educação física - Desporto - Dança</i>		
<i>Arte oratória - Arte dramática</i>		
<i>Fotografia</i>		
<i>Cruz Vermelha</i>		

O estudo de cada aluno antes da sua entrada em New Trier

Dissemos que a escola recolhe informações pormenorizadas sobre os futuros alunos a fim de que eles possam encontrar imediatamente lugar, pelo menos provisório, no sistema complexo que acabámos de descrever. Vamos ver como se procede.

A maior parte dos estudantes que frequentam a New Trier provêm de seis escolas públicas e de sete escolas confessionais dos arredores.

A partir do mês de Janeiro, os serviços de *testing* da High School fornecem às seis escolas públicas testes de inteligência, de aptidão e de conhecimentos a ministrar às crianças que manifestavam a intenção de se matricularem na New Trier. Os alunos das escolas confessionais virão a Winnetka, em Abril, para fazerem provas similares.

Cada escola deverá ainda preencher uma ficha individual que contém, além dos resultados dos testes mencionados, um quadro que resume a apreciação geral dos professores¹:

APRECIACÕES GLOBAIS

Inteligência	1	2	3	4	5
Aplicação	1	2	3	4	5
Sentido das responsabilidades	1	2	3	4	5
Comportamento na escola	1	2	3	4	5
Qualidades de chefe	1	2	3	4	5

APRECIACÕES ESPECIAIS

Inglês	1	2	3	4	5
Matemática:					
Raciocínio	1	2	3	4	5
Conhecimentos de base	1	2	3	4	5
«Social Studies»	1	2	3	4	5
Ciências	1	2	3	4	5

¹ Cf. NEW TRIER HIGH SCHOOL, *Test an Personal Data Card.*

Código: 1 = superior; 2 = acima da média; 3 = médio; 4 = abaixo da média; 5 = fraco. Enquadra-se o algarismo que convém.

Após a administração dos testes, o director de serviço do *counseling* de New Trier encontra-se com os conselheiros das diferentes escolas inferiores para recolher elementos complementares que dizem respeito não apenas à instrução, mas também à história, família e saúde dos futuros estudantes.

Em seguida, entrevista-se cada um deles na sua escola. A troca de impressões diz respeito aos talentos (arte, desporto, passa-tempo) e aos gostos (disciplinas preferidas); depois à família, procurando identificar os problemas eventuais. Uma pergunta-choque termina a entrevista.

No mês de Maio, convidam-se pais e alunos — que já tiveram a oportunidade de estudar os programas dos cursos — para uma reunião de informação na High School. Recolhem-se as inscrições nos diversos cursos.

A direcção da New Trier fica assim com elementos bastantes para constituir grupos de cerca de 30 alunos que, durante todos os seus estudos, serão confiados ao mesmo conselheiro pedagógico e terão um encontro com ele todas as manhãs. Nestes grupos, procura-se sistematicamente a heterogeneidade, tanto no que se refere às aptidões como as origens socio-económicas. Todavia os sexos separam-se, sendo o conselheiro sempre do mesmo sexo que os estudantes. Além disso, faz-se o possível para dar uma personalidade equilibrada a cada grupo, evitando de modo especial que se juntem muitos adolescentes com «problemas».

Ao iniciar o ano, estão organizados não só os cursos e os horários como ainda cada estudante sabe exactamente o que vai fazer. Sabe, por exemplo, que se deve juntar ao grupo dos médios para determinada disciplina, ao grupo superior para outra. Sabe que o conselheiro está às ordens. Sabe ainda que livros, cadernos, equipamento desportivo deverá comprar e já foram informados do preço exacto de cada um desses artigos.

Terminaremos estas indicações respeitantes ao início do ano escolar, assinalando que, cerca de 15 dias depois do começo das aulas, os pais serão convidados para uma reunião em que irão encontrar-se com o conselheiro dos seus filhos e em que serão informados do porquê das classificações dos diferentes níveis, do significado das actividades extra-escolares, etc.

Na semana seguinte, pais e mães irão à escola com os respectivos filhos e assistirão a um dia completo de aulas ao lado deles (neste caso o dia divide-se em tarde e noite). É fácil concluir como uma experiência destas irá facilitar a compreensão entre a escola e a família.

Tendo começado os estudos secundários superiores nestas condições óptimas, o estudante será seguido, dia após dia, pelo seu conselheiro.

Os conselheiros pedagógicos

Graças aos seus conselheiros pedagógicos, a New Trier High School conseguiu essa coisa extraordinária que consiste não apenas em guiar eficazmente cada um dos seus 3740 estudantes no dédalo dos programas mas ainda em ajudá-los tanto no plano psicológico como no plano social e médico.

O *counseling* é assegurado por:

126 conselheiros pedagógicos em *part-time*, professores que consagram 1/5 do seu tempo a essa missão;

10 conselheiros *full-time* que dirigem os precedentes;

6 psicólogos *full-time*.

No decurso da reunião quotidiana da manhã, cada grupo de 30 estudantes cuja constituição já referimos vai ter com o conselheiro para discutir problemas de disciplina, de profissão e, em geral, de tudo o que poderá vir a interessá-los pela vida

fora. A atmosfera descontraída desses encontros constitui uma excelente transição para o trabalho árduo que em breve irão começar.

O estudante poderá ainda consultar individualmente o conselheiro, sempre que o desejar.

Além disso, o conselheiro, durante o ano, fará uma visita pelo menos à família de cada aluno pertencente ao seu grupo.

É evidente que tão delicada missão não se improvisa. Reclama homens abertos aos problemas da juventude e especialmente formados para a ajudar.

Por ocasião do recrutamento dos professores, o conselho de administração da escola dá uma particularíssima importância às qualidades intelectuais e morais indispensáveis à orientação dos estudantes e dá a prioridade àqueles que as possuem.

Durante o seu primeiro ano de funções, deverá o conselheiro assistir a 50 reuniões de formação de cerca de meia-hora cada uma e, após cada ciclo de 4 anos, será submetido a um novo treino.

A preparação de base desenrola-se da seguinte maneira:

1. Antes do começo das aulas:

Consagra-se um certo número de sessões ao estudo da administração da escola. Em seguida, o conselheiro receberá comunicação de todas as informações recolhidas sobre os seus futuros alunos.

Enfim, alguns dias antes do fim das férias, o director de estudos do primeiro ano reúne os «freshman helpers», isto é, estudantes do último ano que irão ajudar o conselheiro na sua tarefa. Durante as 9 primeiras semanas, o assistente participará em todas as reuniões da manhã. Durante as nove semanas seguintes, estará presente duas vezes por semana. Depois disso, apenas intervirá se solicitado pelo conselheiro.

2. Durante o ano escolar

Sob a presidência dum director especializado, a formação prossegue por ocasião das palestras familiares feitas pelos directores dos departamentos (Inglês, Matemática, Ciências, etc.), pelos chefes dos diferentes serviços (directores do *testing*, psicólogo, bibliotecário, médico-assistente social, delegado da associação dos pais) e pelos adultos responsáveis dos clubes de estudantes.

O director dos conselheiros convoca ainda certas reuniões especiais para encarar problemas novos que dizem respeito à admissão de alunos, revisão de programas, etc.

Em Abril, uma conferência final das actividades permitirá fazer o balanço do trabalho do ano que vai terminar e estabelecer as primeiras previsões para o ano seguinte.

BIBLIOGRAFIA

- AGAZZI, A. — *Les aspects pédagogiques des examens*, Strasbourg, Conseil de l'Europe, 1967.
- BACHER, F. — La normalisation de la notation, in *Docimologie et Education*, numéro especial da revista «Les Sciences de l'éducation», n.º 2-3, 1969, 131-156.
- BACHER, F. — L'évaluation des résultats scolaires au niveau de l'école moyenne, in *Le Travail humain*, 1965, 28, 219-230.
- BACHER, F. — La docimologie, in *Traité de psychologie appliquée*, VI, Paris, P.U.F., 1973, pp. 27-86.
- BAZIN, R. — Les Français d'aujourd'hui et leurs examens, in *Education et Gestion*, 4, 1970, 3-10.
- BLOCK, J. — *Mastery Learning*, New York, Holt, Rinehart and Winston, 1970.
- BLOOM, B., HASTING and MADAUS — *Formative and Summative Evaluation of Student Learning*, New York, Mc Graw-Hill, 1970.
- BLOOM, B. S. — *Time and Learning*, comunicação ao 81.º congresso da American Psychological Association, Montréal, 1973.
- BONBOIR, A. — *La docimologie*, Paris, P.U.F., 1972.
- BONNARDEL, R. — Application de la méthode d'analyse factorielle de Thurstone à l'étude de la notation des copies d'examens, in *Le Travail humain*, VIII, 1946, 130-139.
- BRITTON, J. — Experimental Marking of English Composition Written by Fifteen-Year-Olds, in *Educational Review*, Birmingham, vol. 16.1, 1963, 17-23.
- BRITTON, J., MARTIN, N. e ROSEN, H. — *Multiple Marking of English Composition, An Account of an Experiment* London, H. M.S.O., 1966.
- BRUCE, G. — *Secondary School Examinations. Facts and Commentary*, Oxford, Pergamon Press, 1969.
- BRUNELLE, L. — *Pourquoi des examens?* Paris, Société des Editions Rationalistes, 1968.
- CARDINET, J. — *L'adaptations des tests aux finalités de l'évaluation*, Neuchâtel, Institut Romand de Recherche Pédagogique et de Documentation Pédagogique, 1972, 31 p.

CHOPPIN, B. e PURVES, A. — A comparison of open-ended and multiple choice items dealing with literary understanding, in *Research in the Teaching of English*, 3, 1, 1969, 15-24.

COMBER, L. C. and KEEVES, J. — *Science Education in Nineteen Countries*, I.E.A., Stockholm, Malmqvist, 1973.

DEMANGEON, M. e LARCEBEAU, S. — *Une expérience de correction multiple*, in BINOP, 1958, 14, 131-156.

Docimologie et Education, número especial da revista *Les Sciences de l'Education*, 2-3, 1969, 166 p.

EDGEWORTH, F. V. — The Statistics of Examinations, in *Journal of the Royal Stat. Society*, 1888, 51, 599-635.

ELLEY, B. E. and D. LIVINGSTONE, I. — *External Examinations and Internal Assessments. Alternative Plans for Reform*, Wellington, New Zealand Council for Educational Research, 1972.

Examen des examens, número especial dos Cahiers de Pédagogie, 92, Setembro de 1970.

Examinations Bulletins, Londres, H.M.S.O.

Nº 1. *The Certificate of secondary education: some suggestions for teachers and examiners*, 1963.

Nº 2. *The C. S. E.: Experimental examinations — Mathematics*, 1964.

Nº 3. *The C.S.E.: An introduction to some techniques of examining*.

Nº 4. *The C.S.E.: An introduction to objective-types examinations*, 1964.

Nº 5. *The C.S.E.: School-based examinations*, 1965.

Nº 6. *The C.S.E.: Experimental examinations: Technical drawing*, 1965.

Nº 7. *The C.S.E.: Experimental examinations — Mathematics 2*, 1965.

Nº 8. *The C.S.E.: Experimental examinations: Science*, 1965.

Nº 9. *The C.S.E.: Trial examinations: Home economics*, 1966.

Nº 10. *The C.S.E.: Experimental examinations: Music*, 1966.

Nº 11. *The C.S.E.: Trial examinations — Oral English*, 1966.

Nº 12. *Multiple marking of English compositions*, 1966.

Nº 13. *The C.S.E.: Trial examinations: Handicraft*, 1966.

Nº 14. *The C.S.E.: Trial Examinations — Geography*, 1966.

- N° 15. *Teachers experience of school based examining (English and Physics)*, 1967.
- N° 16. *The C.S.E.: Trial examinations —Written English*, 1967.
- N° 17. *The C.S.E.: Trial examinations — Religious knowledge*, 1967.
- N° 18. *The C.S.E.: The place of the personal topic — History*, 1968.
- N° 19. *C.S.E.: Practical work in science*, 1969.
- N° 20. *C.S.E.: A Group Study Approach to Research and Development*, Londres, Evans-Methuen, 1970.
- FISCHER, H. — Wechselwirkungen zwischen Unterrichtszielen, Didaktik und Prüfungen, in *Eidgenössischen Technischen Hochschulen Bulletin* (Zürich), août 1970, 9-14.
- FRENCH, J. W. — *Schools of Thought in Judging Excellence in English Themes*, Princeton, E.T.S., 1961.
- HARTOG, P. and RHODES, E. C. — *An Examination of Examinations*, London, McMillan, 1936.
- HARTOG, P. — *The Marking of English Essays*, London, McMillan, 1941.
- HINTON, E. M. — *An Analytical Study of the Qualities of Style and Rhetoric Found in English Compositions*, New York, 1940.
- HOTYAT, F. — *Les examens*, Paris, Bourrelrier, 1962.
- INGENKAMP, K. — *Die Fragwürdigkeit der Zensurengebung*, Weinheim, Beltz, 1971.
- LAUGIER, H. et SCHREIDER, E. — Recherche docimologique sur un examen de l'enseignement supérieur, in *Biotypologie*, 1958, 1958, 19, n° 2, 61-72.
- LAUWERS, J. A. et SCANLON, D. G. Ed. — *Examinations*, The World Year Book of Education, 1969, London, Evans, 1969.
- LLOYD, W. A. — Les examens en Angleterre, in *Revue Française de Pédagogie*, Janvier 1968.
- MATHER, D., FRANCE, N. et SARE, G. — *The C.S.E., A Handbook for Moderators*, London, Collins, 1965.
- MCINTOSH, D., WALKER, D. and MCKAY, D. — *The Scaling of Teachers Marks and Estimates*, Edinburgh, Oliver and Boyd, 1962, 2^e éd.
- MONTGOMERY, R. J. — *Examinations. An Account of their Evolution as Administrative Devices in England*, Londres, Longmans, 1965.

- JOINT MATRICULATION BOARD — *The Marking of Scripts at Advanced Level History*, Universities of Manchester, Liverpool, Leeds, Sheffield and Birmingham, 1964.
- OTTER, H. S. — *A Functional Language Examination*, Oxford Univ. Press, 1968.
- PASSERON, J. C. — Sociologie des examens, in *Education et Gestion*, 1970, 2, 6-16.
- PEDLEY, F. H. — *A Parent's Guide to Examinations*, Oxford, Pergamon Press, 1964.
- PIDGEON, D. et YATES, A. — *An Introduction to Educational Measurement*, Londres, Routledge et Kegan Paul, 1968.
- PIÉRON, H. — *Examens et docimologie*, Paris, P.U.F., 1963.
- PIÉRON, H., REUCHLIN, M. et BACHER, F. — *Une recherche expérimentale de docimologie sur les examens oraux de physique au niveau du baccalauréat de mathématiques*, in *Biotypologie*, 1962, 23, 48-73.
- PIOBETTA, J. B. — *Examens et Concours*, Paris, P.U.F., 1943.
- PURVES, A. C. — *Literature Education in Ten Countries*, I.E.A., Stockholm, Malmqvist, 1973.
- REMONDINO, C. — Recherche sur les systèmes numériques d'évaluation scolaire, in *Le Travail humain*, 1965, 18, 3-4, 263-265.
- Reports of the Secondary School Examinations Council*, Londres, H.M.S.O., 1974, (1st) — 1964 (8st).
- REUCHLIN, M. — *L'orientation pendant la période scolaire*, Strasbourg, Conseil de l'Europe, C.C.C., 1964.
- REUCHLIN, M. et BACHER, F. — L'appréciation des élèves par leurs professeurs, in *Revue française de Pédagogie*, 1968, 2, 19-25.
- ROLLER, S. — L'évaluation du travail pédagogique, in *Educateur et bulletin corporatif* (Montreux), 1970, 36, 694-696.
- ROT, N. et BUJAS, Z. — Les distributions de notes scolaires comparées aux distributions des résultats obtenus dans les tests de connaissances, in *Le Travail humain*, 1959, 22, 19-26.
- THORNDIKE, R. L. — Marks and Marking Systems, in R. L. EBEL, *Encyclopedia of Educational Research*, Londres, McMillan, 1969, pp. 759-766.
- THORNDIKE, R. L. — *Reading Comprehension in Fifteen Countries*, I. E. A., Stockholm, Malmqvist, 1973.

- Symposium sur la docimologie, XIII^e Congrès de l'Association Internationale de Psychologie (Rome, 1958), in *Le Travail humain*, XXII, 1-2, janvier-juin 1959.
- TYLER, R., GAGNE, R. et SCRIVEN, N. — *Perspectives of Curriculum Evaluation*, AERA Monograph series on curriculum eval., N° 1, Chicago, Rand McNally, 1967.
- VALENTINE, C. W. — *The Reliability of Examinations*, University of London Press, 1932.
- VERNON, P. — *Secondary School Selection*, Londres, Methuen, 1957.
- WALKER, A. S. — *Pupils' School Records*, Newnes, Educ. Publ., 1955.
- WENDLER, J. — *Standartarbeiten, Verfahren zur Objektivierung der Notengebung*, Weinheim, J. Beltz, 1969.
- WISEMAN, S. — The Marking of English Composition in Grammar School Selection, in *British Journal of Educational Psychology*, XIX, 1949, 200-209.
- WISEMAN, S. — *Examinations and English Education*, Manchester, University Press, 1961.
- WRIGLEY, J. — The Relative Efficiency of Intelligence and Attainment Tests as Predictors of Success in Grammar Schools, in *British Journal of Educational Psychology*, 25, 1955, 107-116.
- YATES, A. and PIDGEON, D. — *Admission to Grammar Schools*, London N.F.R.E., 1957.

ÍNDICE ANALÍTICA

INTRODUÇÃO	7
----------------------	---

PRIMEIRA PARTE

DEFINIÇÕES

I. Docimologia, docimástica e doxologia	13
II. Exames e concursos:	
Observação e avaliação contínuas	14
Exames internos e exames externos	16
III. Medida e avaliação	16
IV. Notas e <i>scores</i>	17

SEGUNDA PARTE

A ACUSAÇÃO E A DEFESA

CAPÍTULO 1.º — Crítica aos exames	21
1. Corpos estranhos na educação, ao serviço duma pedagogia ultrapassada	21
2. Ansiedade e <i>stress</i>	22
3. Desigualdade — injustiça	23
4. O fracasso, gerador de fracassos	30
5. Ruptura entre ensino e exames	31
6. Desacordo entre os que corrigem	32
a) Composição francesa	33
b) Matemática	35
c) Medicina	35
d) Diversos	36
e) Nos interrogatórios orais, ainda mais discordâncias	37
f) Quantas pessoas a corrigir para estabilizar a nota?	37

7. Infidelidade no mesmo apreciador	39
Um esquema para continuar a investigação	40
8. Estereótipos e efeitos de halo	41
9. Efeitos de ordem de correcção	46
10. Falta de validade	48
11. Um instrumento de imobilismo social	48
a) Efeitos irreversíveis do certificado escolar	49
b) Os exames não são socialmente neutros	49
12. Defeitos de muitas experiências docimológicas	53
13. Outras críticas	53
CAPÍTULO 2.º — Defesa da nota subjectiva e do exame	56
1. A medida rigorosa é talvez impossível	57
2. Os professores julgam bem os alunos	59
3. Validade limitada mas real dos exames tradicionais	61
4. Endurecer para a vida	62
5. Situar-se em relação aos outros	62
6. Larga síntese e integração dos conhecimentos	62
7. O exame externo controla o professor	63
8. O exame externo, <i>feedback</i> para o professor	63

TERCEIRA PARTE

CONSTRUÇÃO DO EXAME

As grandes fases — Vista de conjunto	67
CAPÍTULO 1.º — O objecto e os objectivos	68
I. O objecto	68
A. O prognóstico	69
1. Testes de maturidade específica (<i>readiness</i>)	69
2. Verificação dos conhecimentos-chave ou noções críticas	70
3. Ensaio	70

B. O inventário	70
C. O diagnóstico	71
II. Os objectivos	73
A. Os objectivos gerais	74
1. Os objectivos cognitivos	75
a) A taxonomia de Bloom	75
b) O modelo de Guilford	79
2. Os objectivos afectivos	83
B. Os objectivos especiais	85
C. Os objectivos operacionais	93
CAPÍTULO 2.º — Redacção das perguntas	97
I. Observações gerais	97
A. Perguntas compreensíveis	99
B. Ter em conta o nível de informação	99
C. Ensaiar ou pretestar as perguntas	100
D. Cálculo da facilidade das perguntas	100
E. Cálculo da eficácia — Poder discriminativo	101
1. Método simples	101
2. Método mais apurado	101
II. Respostas abertas ou fechadas?	105
A. Respostas abertas	106
B. Respostas fechadas — Perguntas de escolha múltipla	107
1. Utilidade	107
2. Organizar uma provisão de perguntas	108
3. Explorar a gama das possibilidades lógicas	108
a) Perguntas de complemento simples	109
b) Associação simples	110
c) Associação composta	110
d) Associação com exclusão de termo	111
e) Análise de relação de causa a efeito	111
f) Análise de observação	112
g) Comparações quantitativas	114
h) Relações	114
i) Complementos agrupados	115

4.	Cálculo da eficácia dos diversores	115
5.	Críticas e refutação parcial	116
	a) Uma objectividade enganadora	116
	b) Escolhas «correctas» contestáveis	117
	c) Uma lotaria	117
	d) Acrobacia mental.	118
	e) Inconvenientes incertos	120
	C. A modo de conclusão: um compromisso	120
III.	Subjectividade — Objectividade	121
	A. Teoria	121
	B. Alguns exemplos	125
	1. O teste de «closure»	125
	2. Teste de compreensão da leitura.	126
	3. Formas de <i>items</i> para a subtracção	128
	4. Exemplo do sistema de geração de <i>items</i> para a Matemática nova no início da escola primária	129
	CAPÍTULO 3.º — A classificação	135
	I. Um preâmbulo indispensável: a curva de Gauss	135
	A. A curva de Gauss, imagem da probabilidade.	135
	B. A curva de Gauss, imagem do ensino não individualizado	137
	C. O desvio-padrão ou sigma, índice precioso	138
	1. Significado	138
	2. Avaliação rápida	139
	D. A concentração dos resultados à volta da média	142
	E. A curva de Gauss desejada pelos professores	143
	F. Como saber se uma distribuição é normal?	143
	II. A classificação subjectiva: a escala de avaliação	147
	A. Introdução	147
	B. Natureza e faltas das escalas de cotação	148
	C. Utilidade	150

D.	Construção	150
1.	Quantos graus?	151
2.	Definir o objecto da avaliação	152
E.	Utilização	157
1.	Quantos alunos por escalão?	157
	— Aluno comparado a si mesmo	157
	— Alunos comparados entre si	159
2.	Lutar contra a contaminação e contra a tendência central	160
F.	Como sintetizar as avaliações?	161
G.	Um caso particular: a classificação da composição francesa	165
1.	Quatro métodos de avaliação	166
a)	O método da impressão geral	166
b)	As escalas de espécimes	168
c)	O método analítico	169
	Que qualidades observar?	170
	Exemplos	172
d)	O método de contagem das frequências	178
2.	Vários assuntos à escolha?	179
3.	Conclusão	181
III.	A classificação objectiva	183
IV.	A aferição ou medida da posição relativa	184
IV-1.	Aferição dos testes normativos	184
A.	Centilagem	185
B.	As notas standard ou notas Z	189
C.	Escala de cinco classes	192
D.	escala de nove classes (<i>stanines</i>)	193
IV-2.	A aferição em relação ao objectivo	194
CAPÍTULO 4.º — Controlo da fidelidade do exame		197
1.	Evitar toda a ambiguidade nas perguntas	197
2.	Perguntas em número suficiente	198
3.	Um controlo matemático	198
a)	O método pares-ímpares	198
b)	Duas formas paralelas	199
4.	Repetição da classificação	200

CAPÍTULO 5.º — Controlo da validade	201
I. Validade do conteúdo	201
II. Validade da previsão.	204

QUARTA PARTE

OS PROCESSOS DE MODERAÇÃO

CAPÍTULO 1.º — Posição do problema	209
1. Definição	209
2. Moderar não é arremeter	211
3. Moderação voluntária ou imposta?	211
4. A moderação começa no início do ano escolar	212
5. Não há possibilidade de comparação sem fidelidade elevada	213
6. Poderemos fiar-nos nos testes?	213
CAPÍTULO 2.º — Alguns sistemas de moderação dos exames	216
I. Por referência a um teste	216
A. A fórmula mais liberal: o sistema sueco de moderação por disciplina a partir de testes de conhecimentos	216
B. Sistema de moderação imposto por disciplina a partir dum teste de conhecimentos	219
C. Um sistema de selecção a partir dum teste de inteligência	220
II. Moderação por recurso a uma banca de <i>items</i>	223
III. Processo de equilíbrio	224
Um sistema completo de moderação em Inglaterra	224
A. Preliminares	225
B. Os professores classificam os exames.	226
C. Correção pelos moderadores	226
1.º Controlo da severidade	229
2.º Controlo da discriminação	230
3.º Controlo da conformidade	230
D. Correção das amostras restantes e controlos	230

E. Como ajustar notas discordantes?	233
1.º Ajustamento da mediana	233
2.º Ajustamento da mediana e do desvio-padrão.	233
F. Nota do fim do ano	236
Trabalho do ano + trabalhos práticos + teste	236

QUINTA PARTE

UMA PEDAGOGIA DA DOCÊNCIA

(Mastery Learning)

O perigoso mito da curva de Gauss	245
CAPÍTULO 1.º — Evolução da curva dos conhecimentos	246
I. Curva das aptidões	246
II. Curva dos conhecimentos	248
CAPÍTULO 2.º — Uma pedagogia da curva em J	250
CAPÍTULO 3.º — A teoria da avaliação formativa	253
I. Escalonar a ascensão do saber	255
II. Guiar o aluno	261
III. A relação tempo-aprendizagem	262
IV. O sistema de ensino posto em causa	265
CAPÍTULO 4.º — A teoria da avaliação somativa	268

CONCLUSÕES E RECOMENDAÇÕES

CONCLUSÕES E RECOMENDAÇÕES	273
--------------------------------------	-----

ANEXOS

I.	Estudo comparado duma pergunta de exame apresentada segundo o método tradicional e segundo o método com perguntas de escolha múltipla	279
	— Método tradicional	279
	— Método por escolha múltipla	280
II.	Exemplo de perguntas para uma composição de língua materna	
	— Inglaterra, C. G. E., 1967 — Fim do ensino secundário . . .	286
	— Perguntas	286
	Prova I	286
	Prova II	287
	— Indicações para a correcção	289
	Prova I	289
	Prova II	292
III.	Exemplo de ensino semi-individualizado	296
	— A New Trier Township High School.	296
	— A individualização dos programas	297
	— A individualização do ensino.	301
	— O estudo de cada aluno antes da sua entrada em New Trier	307
	— Os conselheiros pedagógicos	309
	BIBLIOGRAFIA	313

ÍNDICE

INTRODUÇÃO	7
PRIMEIRA PARTE — <i>Definições</i>	11
SEGUNDA PARTE — <i>A acusação e a defesa</i>	19
Capítulo 1: Crítica dos exames	21
Capítulo 2: Defesa da nota subjectiva e do exame	56
TERCEIRA PARTE — <i>Organização do exame</i>	65
Capítulo 1: O objecto e os objectivos	68
I. Objecto	68
II. Os objectivos	73
Capítulo 2: A redacção das perguntas	97
I. Observações gerais	97
II. Respostas abertas ou fechadas?	105
III. Subjectividade-objectividade	121
Capítulo 3: A classificação	135
I. Um preâmbulo indispensável. A curva de Gauss	135
II. A classificação subjectiva. Escala de avaliação	147
III. A classificação objectiva.	183
IV. A aferição ou medida da posição relativa.	184
Capítulo 4: O controlo da fidelidade	197
Capítulo 5: O controlo da validade	201
QUARTA PARTE — <i>Os processos de moderação</i>	207
Capítulo 1: Posição do problema	209
Capítulo 2: Alguns sistemas de moderação	216
I. Por referência a um ou vários testes.	216

II. Moderação por recurso a uma banca de <i>items</i>	223
III. Processo de equilíbrio	224
Em Inglaterra. Um sistema de moderação completo	224
QUINTA PARTE — <i>Uma pedagogia da docência (Mastery learning)</i>	243
O mito perigoso da curva de Gauss	245
Capítulo 1: Evolução da curva dos conhecimentos.	346
Capítulo 2: Uma pedagogia da curva em J	250
Capítulo 3: A teoria da avaliação formativa	253
Capítulo 4: A teoria da avaliação somativa	268
CONCLUSÕES E RECOMENDAÇÕES	261
ANEXOS	277
I. Estudo comparado duma pergunta de exame	279
II. Exemplo de perguntas para uma composição em língua materna	286
III. Exemplo de ensino semi-individualizado	296
BIBLIOGRAFIA SUCINTA	313
ÍNDICE ANALÍTICO	319



Este livro foi executado para a
Livraria Almedina, nas oficinas da
Gráfica de Coimbra. Janeiro
de 1976

«COLECÇÃO NOVALMEDINA»

Volumes publicados:

- 1 — Avaliação Contínua e Exames — *Gilbert de Landsheere*
- 2 — As Línguas Vivas — *Denis Girard*
- 3 — A Relatividade Educativa — *Arnould Clause*
- 4 — A História e o seu ensino — *Vários professores da Universidade de Québec*
- 5 — Medições e avaliações na aula — *Clarence H. Nelson*