



Título: *A avaliação dos alunos*

Prefácio: Carlos Fiolhais

Autores: Jeffrey Karpicke, Hélder Diniz de Sousa, Leandro S. Almeida

Tradução: Joana Matias

Revisão: Hélder Guegués

Design e paginação: Guidesign

Colecção: Questões-Chave da Educação

Edição: Fundação Francisco Manuel dos Santos

1.ª edição: Outubro de 2012

© Fundação Francisco Manuel dos Santos, 2012

Impressão: Guide Artes Gráficas, Lda.

ISBN: 978-989-8424-61-7

Depósito Legal n.º: 349768/12

FUNDAÇÃO FRANCISCO MANUEL DOS SANTOS

Rua Tierno Galvan, Torre 3, 9ºJ

1070-274 Lisboa

Telf: 21 381 84 47

ffms@ffms.pt

QUESTÕES-CHAVE
DA EDUCAÇÃO

A AVALIAÇÃO DOS ALUNOS

Jeffrey Karpicke
Hélder Diniz de Sousa
Leandro S. Almeida

Prefácio de Carlos Fiolhais

- 7** PREFÁCIO
Carlos Fiolhais
- 17** APRENDIZAGEM COM BASE
NA RECUPERAÇÃO: A RECUPERAÇÃO ACTIVA
PROMOVE UMA APRENDIZAGEM SIGNIFICATIVA
Jeffrey D. Karpicke
- 41** EXAMES NACIONAIS:
INSTRUMENTOS DE REGULAÇÃO DE BOAS
PRÁTICAS DE ENSINO E DE APRENDIZAGEM?
Hélder de Sousa
- 73** AVALIAÇÃO DOS ALUNOS:
COMBINANDO AS RAZÕES E OS MODOS
Leandro S. Almeida

PREFÁCIO

Carlos Fiolhais*

Avaliação é uma palavra que está na ordem do dia em todos os setores na sociedade moderna. Também o está na educação: a começar na avaliação individual dos alunos e a acabar na avaliação do sistema educativo como um todo, passando pela avaliação dos professores e pela avaliação das escolas. Todos estes tipos de avaliação estão relacionados: não se pode, por exemplo, pensar na avaliação de um sistema educativo nacional sem uma adequada avaliação das escolas, dos professores e dos alunos; não se pode obter uma avaliação de escolas sem uma avaliação dos professores e dos alunos; e é difícil conceber a avaliação dos professores se não se atender, entre outros factores, à avaliação que fazem dos seus alunos.

Em Portugal, todos estes tipos de avaliação têm sido alvo de grandes controvérsias: a avaliação do sistema educativo tem sido feita não só informalmente pela opinião pública como, de um ponto de vista formal e com base em indicadores quantitativos, por comparações internacionais como as que são feitas pela União Europeia e pela Organização para a Cooperação e Desenvolvimento Económico (OCDE); a avaliação das escolas é feita, ou deveria ser feita, para lá dos cidadãos que a elas confiam os seus filhos (os *rankings* baseados nos exames

nacionais do final dos ensino básico e secundário representaram uma abertura de um sistema que antes persistia em manter-se opaco), pelos serviços do Ministério da Educação; a avaliação dos professores deu azo entre nós a um gigantesco protesto, que levou mais de cem mil professores a manifestarem-se na rua contra o esquema desmesuradamente burocrático que o ministério queria impor (essa acção levou a um recuo para uma versão “simplificada”, que está a dar os primeiros passos); e a avaliação dos alunos, proporcionada nos últimos tempos por exames nacionais padronizados no 9.º, 11.º e 12.º anos, organizados pelo Gabinete de Avaliação Educacional (GAVE) do Ministério da Educação, ganhou uma atenção e mediatização crescentes, não só pelo papel que a avaliação no final do secundário desempenha no acesso ao ensino superior, mas também porque essas provas dão à sociedade uma medida global do desempenho do sistema de ensino não superior e, portanto, da utilidade da escola.

Este volume, que contém o essencial das intervenções de uma “Conferência sobre Questões-chave” da Educação da Fundação Francisco Manuel dos Santos, centra-se na avaliação dos alunos. Esta é a avaliação mais antiga de todas as que se realizam na educação – nunca houve uma progressão na escola sem uma avaliação, de uma forma ou de outra, das aprendizagens realizadas – e é também aquela que tem o maior número de destinatários – todos os alunos, de uma forma ou de outra, são sujeitos a avaliação ao longo do seu percurso escolar. A avaliação impõe aos alunos patamares de exigência, objectivos e ambições.

Distinguem-se vários tipos de avaliação, mas sumariamente podem dividir-se a avaliação, além da que serve de diagnóstico, na que é formativa, que acompanha o trabalho escolar do aluno, e na que é sumativa, que se destina a fornecer uma classificação final, que, no caso do secundário, poderá servir para ingressar num nível superior de ensino. As duas são

necessárias e as duas complementam-se. Em Portugal, depois do 25 de Abril de 1974, houve uma depreciação da avaliação sumativa, pretensamente em favor da avaliação formativa. De facto, o que aconteceu foi uma diminuição da avaliação no seu conjunto e, portanto, da exigência, dos objectivos e das ambições. A palavra “exames” foi durante alguns anos um termo tabu. Seriam traumáticos para os alunos. Prejudicariam a “verdadeira” aprendizagem. Não teriam grande significado nem importância. Assim, foi possível em Portugal entrar no ensino superior, em que para cada disciplina havia um exame (para muitos, um exagero que não se encontra em todos os sistemas de ensino superior), sem que antes se tivesse passado por um exame digno desse nome. Mas os exames nacionais reapareceram no básico e secundário, revelando a sua utilidade não só para a progressão dos alunos como também como instrumento normalizador e responsabilizador do sistema educativo. Começaram por ser exames no final dos estudos secundários relacionados com a salutar massificação do ensino e a necessidade de introdução de *numerus clausus* no ensino superior. Mas surgiram também, mais recentemente, exames nacionais no final da escolaridade básica em disciplinas consideradas essenciais (usa-se a palavra “estruturantes”) como o Português e a Matemática. Tanto no ensino secundário como no básico a classificação final do aluno resulta de uma média ponderada entre a avaliação sumativa e a avaliação formativa, com clara predominância desta última. Nas referidas disciplinas, instituíram-se também “provas de aferição” no 4.º e no 6.º anos de escolaridade, sendo a diferença destas em relação aos exames o facto de que com elas se pretende fazer um diagnóstico de aprendizagens realizadas e, portanto, medir, tanto quanto possível, o desempenho das escolas e do sistema, em vez de dar classificações que tivessem consequências no percurso educativo dos alunos. Pode dizer-se que a introdução de exames foi,

entre nós, socialmente bem recebida, ficando a ideia geral de que o sistema educativo se tornava com eles mais rigoroso, mais transparente e mais ligado à comunidade.

É decerto de louvar a aposta governamental na organização de uma avaliação dos alunos, uma avaliação correcta, justa e, por isso, credível. Mas, se houve esse cuidado, os resultados não foram, em geral, bons. As notas, em particular nas ditas disciplinas “estruturantes”, não foram as desejáveis, com médias negativas em alguns anos. Não se pode dizer que as provas tenham sido sempre bem preparadas e que sejam comparáveis de uns anos para os outros (algumas oscilações, para baixo e para cima, devem ter essa explicação simples). Provavelmente para evitar que os resultados fossem ainda piores, foi nítida a preocupação de incluir nos enunciados das provas questões que, em alguns casos, eram gritantemente simples (como numa prova intermédia de Física e Química do 9.º ano, preparatória para o exame nacional, em que se indicavam os nomes de todos os planetas e se pedia simplesmente ao aluno que os contasse, de um até oito). Por vezes, ocorreram mesmo erros científicos, algo intolerável numa prova, que não foram prontamente reconhecidos (ficou famosa uma prova de Geometria Descritiva, em que se pedia para desenhar uma certa tangente a uma certa circunferência numa situação em que tal era geometricamente impossível).

Mas, mesmo com todos esses percalços, os exames ficaram instituídos. Nunca desapareceram, porém, as críticas de alguns sectores ao papel redutor dos exames. De facto, avaliar é sempre um acto redutor de uma realidade complexa, mas, tal como na ciência, medir consiste em reduzir a realidade a números e pouco se pode avançar sem isso (é melhor uma medida menos boa do que medida nenhuma!), também na educação será difícil dizer algo substantivo a respeito do sucesso ou insucesso sem recolher medidas minimamente bem-feitas. Associado

à avaliação, há quem critique um ensino “orientado para os exames”, mas o certo é que os bons professores orientam não para os exames mas para os saberes e para as capacidades e as provas, se forem bem-feitas, isto é, feitas por bons professores, não poderão deixar de medir esses saberes e essas capacidades. Os exames, com todas as deficiências que lhes possam ser apontadas, têm servido para confrontar os alunos consigo próprios. Têm servido para que os professores se apercebam melhor das dificuldades nas aprendizagens realizadas pelos seus alunos e, em conformidade, possam ensinar mais e melhor. Têm servido para as escolas se aperfeiçoarem, fornecendo-lhes indicadores preciosos (é evidente que os resultados dos exames têm de ser lidos conjuntamente com outros indicadores, como os socioeconómicos). E têm sido úteis a toda a sociedade, pais e outros encarregados de educação assim como aos contribuintes em geral, para ela adquirir uma noção mais fiel acerca do sistema educativo que mantém.

À semelhança do que acontece na maioria dos países da União Europeia e da OCDE, os exames vieram para ficar em Portugal. O nosso país tornou-se, também neste aspecto, “normal”. É agora natural que os exames sejam não apenas melhorados, como aumentados designadamente ampliando o número de disciplinas em que existem exames no ensino básico e transformando as provas de aferição em provas com resultados que contam, ainda que de início, numa percentagem modesta. Nesta linha, já foram anunciados exames no 4.º e 6.º anos. Tal permitirá a todos os intervenientes – alunos, professores, administração escolar a vários níveis e cidadãos em geral – interiorizar ainda mais uma cultura de avaliação e, portanto, de responsabilização. Há quem veja nos exames o fantasma das reprovações e da discriminação social, mas esse é um problema que a escola pública tem de enfrentar decididamente com novas estratégias e novos meios, como mais e

melhor ajuda aos alunos com maiores dificuldades cognitivas e a diversificação das vias escolares. Se a escola oferece uma preparação para a vida, não será deixando cair barreiras que se preparará melhor para a vida, toda ela repleta de barreiras. O ponto essencial é que a todos seja concedida iguais oportunidades.

Modernamente, as ciências cognitivas vieram revelar um novo factor vantajoso associado à realização de provas. Os conhecimentos ficarão mais consolidados se eles forem chamados, isto é, a inquirição dos alunos sobre aquilo que sabem torna-os mais sábios e a solicitação sobre aquilo de que são capazes torna-os mais capazes. Esta dimensão, que a escola deve usar melhor não apenas em provas finais mas na sala de aula ao longo do tempo lectivo, trouxe, sem dúvida, uma importância acrescida à avaliação no domínio pedagógico em que ela era, por vezes, alvo de críticas muito contundentes.

O norte-americano Jeffrey Karpicke, investigador principal do Laboratório de Cognição e Aprendizagem da Universidade Purdue, nos Estados Unidos, expõe-nos aqui a sua investigação educativa relativa à recuperação do conhecimento que está presente na memória dos alunos. Os dados que traz são científicos, na medida em que resultam de experiências, devidamente controladas, feitas com indivíduos que estão a aprender um certo assunto. As ciências da educação, porque são ciências, ganham com a experimentação científica cuidadosa e evoluem com as conclusões que se podem extrair da informação empírica acumulada.

Para reflectir sobre a realidade nacional, foram convidados Hélder de Sousa, professor do ensino secundário que tem sido director do GAVE (embora compreensivelmente não se exprima aqui como porta-voz oficial, mas sim como estudioso da matéria e bom conhecedor da avaliação educativa em Portugal), que traça um panorama da avaliação recente no nosso país, essencialmente centrado nos exames, e de

Leandro Almeida, professor de Psicologia da Universidade do Minho, onde dirige o Instituto de Educação, que enfatiza que vários tipos de avaliação servem diferentes propósitos, aponta características das boas provas e refere os aspectos psicológicos que não deixam de estar associados à prestação de provas.

A avaliação está e continuará a estar na ordem do dia. Sobre ela, alunos, professores, directores de escolas, especialistas da educação, decisores governamentais e pais terão a sua própria opinião, que, apesar de algum consenso recente sobre as vantagens dos exames, está longe de ser unânime. O prefaciador, ao falar de um assunto que levanta polémica, quis, além de apresentar os trabalhos dos especialistas, juntar a sua opinião, com base na sua experiência docente e de intervenção pedagógica, a um debate público cujo prosseguimento se deseja. A educação é um problema de todos. A Fundação Francisco Manuel dos Santos, empenhada como está no melhor conhecimento da realidade nacional, em particular nessa área decisiva para o futuro colectivo que é a da educação, quer que todos tenham e emitam uma opinião baseada na melhor informação.

*Responsável pelo Programa de Educação da Fundação Francisco Manuel dos Santos

QUESTÕES-CHAVE
DA EDUCAÇÃO

**APRENDIZAGEM
COM BASE NA
RECUPERAÇÃO:
A RECUPERAÇÃO
ACTIVA PROMOVE
UMA APRENDIZAGEM
SIGNIFICATIVA**

Jeffrey Karpicke

APRENDIZAGEM COM BASE NA RECUPERAÇÃO: A RECUPERAÇÃO ACTIVA PROMOVE UMA APRENDIZAGEM SIGNIFICATIVA¹

Jeffrey D. Karpicke

Resumo

Embora não lhe seja conferido o papel central que merece, a recuperação é o processo-chave para perceber a aprendizagem e promovê-la. A aprendizagem é tipicamente identificada com a codificação ou construção de conhecimento, e a recuperação é considerada como uma mera avaliação de uma aprendizagem adquirida em experiência anterior. A aprendizagem baseada na recuperação como é aqui gizada baseia-se no facto de todas as expressões de conhecimento envolverem recuperação e dependerem dessa recuperação de pistas disponíveis num contexto particular. Além disso, cada vez que alguém recupera conhecimento, esse conhecimento é modificado, pois recorrer à recuperação melhora a capacidade de recuperar novo conhecimento

¹ Versão do texto publicado na revista *Current Directions in Psychological Science* (no prelo).

no futuro. Praticar a recuperação não produz uma aprendizagem rotineira e momentânea; produz uma aprendizagem significativa e de longo prazo. No entanto, a prática da recuperação é uma ferramenta que muitos estudantes, cuja consciência metacognitiva desconhecem, não utilizam tantas vezes como deveriam. A recuperação activa é uma estratégia efectiva mas subvalorizada para promover uma aprendizagem significativa.

PALAVRAS-CHAVE — processos de recuperação; aprendizagem; educação; metacognição; aprendizagem significativa.

Introdução

«Se souberes algo, ou tiveres armazenado informação acerca de um acontecimento num passado distante, e não usares essa informação, se nunca pensares nisso, o teu cérebro será funcionalmente equivalente a um outro semelhante, mas que não “contenha” essa informação.»

ENDEL TULVING (1991)

Para entender a aprendizagem é essencial perceber os processos envolvidos em recuperar e reconstruir o conhecimento. Podemos pensar que sabemos algo, que as nossas mentes contêm ou possuem algum conhecimento, mas a única forma de examinar o conhecimento é envolver-se no acto de recuperação. As diferenças na capacidade de recuperar conhecimento poderão não advir não do que está «armazenado» nas nossas mentes, mas sim de diferenças nas pistas de recuperação que estão disponíveis em contextos particulares. Dada a importância fundamental da recuperação para entender a aprendizagem, é surpreendente que os processos de recuperação não tenham recebido mais atenção na pesquisa educacional. Podemos verificar que, ao longo da década passada, muitas obras influentes do National Research

Council acerca da aprendizagem não contém qualquer menção à recuperação (National Research Council, 2000, 2005a, 2005b).

É essencial considerar os processos de recuperação não só porque são centrais no entendimento da aprendizagem, mas também porque o acto de recuperação em si mesmo é uma ferramenta poderosa para melhorar a aprendizagem. Além do mais, praticar a recuperação não produz uma aprendizagem rotineira e momentânea: produz uma aprendizagem significativa e de longo prazo. A ideia de que a recuperação é a peça central para compreender a aprendizagem, juntamente com a relevância da recuperação activa para produzir aprendizagem, é uma abordagem denominada *aprendizagem baseada na recuperação*.

Aprendizagem baseada em esquemas mentais

Pensamos muitas vezes nas nossas mentes como sítios nas nossas cabeças, espaços mentais ou contentores onde armazenamos conhecimentos. Roediger (1980) observou que, durante séculos, a maior parte das metáforas usadas para descrever processos mentais caracterizavam a mente como um espaço físico e o conhecimento como coisas físicas nesse espaço – por exemplo, como bibliotecas cheias de livros ou gabinetes repletos de ficheiros (ver também Moscovitch, 2007). Em educação, a metáfora de um edifício físico é muitas vezes usada para descrever mente e conhecimento. O conhecimento é *construído* por aprendizes que erguem activamente *estruturas* de conhecimento; os investigadores tentam compreender a *arquitectura* da mente; e os orientadores auxiliam os estudantes fornecendo-lhes *andaimes* para a aprendizagem.

Quando as mentes são vistas como locais para armazenar conhecimento, é natural que a atenção se foque nos processos envolvidos na construção de novo conhecimento armazenado. A investigação educacional e

as práticas instrutivas valorizaram a identificação das melhores formas de *codificar* conhecimento e experiências. Os processos de recuperação, envolvidos na utilização de pistas disponíveis para reconstruir activamente o conhecimento, receberam menor atenção. Parece existir a presunção tácita de que a codificação ou construção do conhecimento é, em si mesma, suficiente para assegurar a aprendizagem.

A investigação básica sobre aprendizagem e memória, no entanto, tem enfatizado que a recuperação deve ser considerada em qualquer análise da aprendizagem. Em parte, porque as pessoas não armazenam cópias estáticas e exactas de experiências que são reproduzidas literalmente na recuperação. Por outro lado, o conhecimento é activamente reconstruído com base no contexto e pistas de recuperação disponíveis no presente (Bartlett, 1932; Neisser, 1967; ver também Moscovitch, 2007; Roediger, 2000). A natureza reconstrutiva da mente é revelada nos erros sistemáticos que as pessoas cometem na recuperação do conhecimento, erros esses que aparelhos de gravação decerto não fariam. Como o passado nunca ocorre duas vezes da mesma maneira, um armazém mental de cópias de experiências seria de pouca utilidade. Pelo contrário, as pessoas têm a capacidade de utilizar o passado para responder às necessidades do presente, reconstruindo o conhecimento, em vez de o reproduzir exactamente.

O que as pessoas expressam quando reconstroem conhecimento depende das pistas de recuperação que estão disponíveis num contexto particular. Em última análise, a reconstrução do conhecimento depende do valor diagnóstico dessas pistas, o nível a que estas ajudam as pessoas a recuperar informação-chave, em particular excluindo candidatos competidores (Nairne, 2002; Raaijmakers & Shiffrin, 1981). Podemos querer examinar o que uma pessoa construiu ou armazenou na sua mente, mas é impossível aceder directamente aos conteúdos armazenados *per si*.

Apenas podemos examinar o que uma pessoa reconstrói dadas as pistas disponíveis e o contexto (Tulving & Pearlstone, 1966; Roediger & Guynn, 1996; Roediger, 2000). Assim, é essencial considerar os processos de recuperação em qualquer análise da aprendizagem.

A segunda razão crucial para tomar como importante a recuperação na aprendizagem é a alteração da aprendizagem pelo acto da recuperação. Cada vez que alguém recupera conhecimento, esse conhecimento é modificado pois a recuperação melhora a capacidade que cada um tem de voltar a recuperar essa informação no futuro (Karpicke & Roediger, 2007, 2008; Karpicke & Zaromb, 2010). Esta é uma característica da aprendizagem funcional e sistema de memória. As nossas mentes são sensíveis à probabilidade de voltarmos a recuperar conhecimento no futuro, e, se recuperarmos algo no presente, será provável que venhamos a ter de o recuperar de novo. O processo de recuperação em si mesmo altera o conhecimento em antecipação das necessidades que possamos encontrar no futuro. A recuperação não é, portanto, apenas uma ferramenta para avaliar a aprendizagem, mas também um meio para aperfeiçoar essa aprendizagem (Roediger & Karpicke, 2006a).

A recuperação repetida promove a aprendizagem a longo prazo

Imagine o leitor que está a estudar com vista a um futuro exame. Depois de ler as suas notas e/ou livro de estudo uma vez, o que faria de seguida? Tem três opções. Pode:

- (a) voltar atrás e estudar de novo toda a matéria ou parte dela,
- (b) tentar lembrar-se da matéria, sem voltar a estudar, ou
- (c) fazer outra coisa.

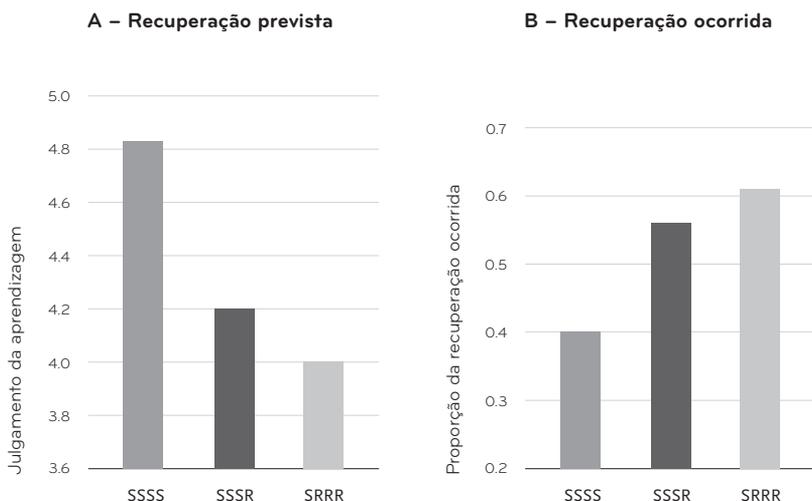
O que escolheria?

Karpicke, Butler e Roediger (2009) puseram esta questão a um vasto grupo de estudantes universitários. A maior parte dos estudantes (57 %) responderam que iriam reler as notas ou livros, mas 21 % disseram que fariam outra coisa. Apenas 18 % disseram que iriam tentar lembrar-se da matéria depois de a ler. A decisão de ler repetidamente faz sentido se identificarmos a aprendizagem com processos de codificação e construção do conhecimento e considerarmos a recuperação apenas como uma avaliação prévia da aprendizagem. Mais estudo (mais codificação e construção de conhecimento) deveria produzir mais aprendizagem, ao passo que a recuperação iria medi-la mas não produzi-la.

Ficariam os estudantes mais bem preparados com a leitura repetida do que recorrendo à recuperação? Roediger e Karpicke (2006b) conduziram duas experiências que correspondiam quase inteiramente à questão colocada no inquérito. Numa experiência, os estudantes leram textos educacionais e relembrou-nos sob três condições. Um grupo de estudantes passou o tempo a estudar repetidamente um texto em quatro períodos de estudo (classificados como SSSS). Um segundo grupo leu um texto em três períodos de estudo e depois tentou lembrá-lo num período de recuperação (SSSR), no qual os estudantes escreveram tudo aquilo de que se conseguiam lembrar sobre o texto. Um terceiro grupo leu o texto num período de estudo e recapitulou-o em três períodos de recuperação repetidos (SRRR). Os estudantes não releeram o texto nem receberam qualquer *feedback* depois de cada período de recapitulação; apenas praticaram activamente a recuperação da matéria. No final da fase de aprendizagem, os estudantes fizeram um *juízo de aprendizagem*, uma previsão de quão bem se lembrariam no futuro da matéria estudada. Então, uma semana depois, os estudantes recapitularam a matéria outra vez para verem que matéria tinham, de facto, retido a longo prazo.

A figura 1 mostra os julgamentos de aprendizagem dos estudantes. Quantas mais vezes os estudantes liam repetidamente a matéria, mais achavam que a tinham apreendido. No entanto, a figura 1B mostra que a aprendizagem real dos estudantes exibia um padrão oposto. Quantas mais vezes os estudantes praticavam a recuperação activamente, melhor retinham o conhecimento a longo prazo. Os estudantes passaram o mesmo período de tempo a estudar a matéria nas três condições e os estudantes que recuperaram a matéria repetidamente apenas a recapitularam, não tendo voltado a estudar os textos. No entanto, a recuperação activa produziu melhor retenção de informação a longo prazo (para um debate sobre a consciência metacognitiva de efeitos práticos da recuperação, ver Karpicke & Grimaldi, 2011).

Figura 1. Recuperação ocorrida (B) depois de estudar um texto em quatro períodos de estudo (condição SSSS), ler um texto em três períodos de estudo e recuperá-lo num período de recordação (condição SSSR), ou ler um texto num período de estudo e recuperá-lo repetidamente em três períodos de recuperação (condição SRRR). Os juízos de aprendizagem (A) foram feitos numa escalas de sete pontos, na qual 7 correspondia à convicção dos alunos de que se lembrariam muito bem do material. Os dados apresentados são adaptados da Experiência 2 de Roediger e Karpicke (2006b). O padrão dos julgamentos metacognitivos dos estudantes (recuperação prevista) foi exactamente o oposto do padrão de retenção a longo prazo dos estudantes.



Voltando ao inquérito sobre as estratégias de aprendizagem dos estudantes (Karpicke *et al.*, 2009), poder-se-á pensar que os resultados seriam diferentes se mudássemos a forma como a pergunta foi feita. A saber, os estudantes poderiam optar pela recuperação activa se pudessem reler depois de tentarem a recuperação. Karpicke *et al.* fizeram aos alunos exactamente a mesma pergunta descrita antes, mas com a opção (b) mudada para «tentar recapitular a matéria e depois voltar atrás e estudar de novo o texto». As opções dos alunos modificaram-se mas não tão acentuadamente como seria de esperar. Agora, 42 % dos alunos disseram que poriam em prática a recuperação e voltariam então a ler. No entanto, 41 % dos alunos continuaram a dizer que apenas voltariam a ler e 17 % disseram que fariam algo diferente. Por outras palavras, 58 % dos alunos indicaram que não praticariam a recuperação activa mesmo que tivessem a hipótese de reler.

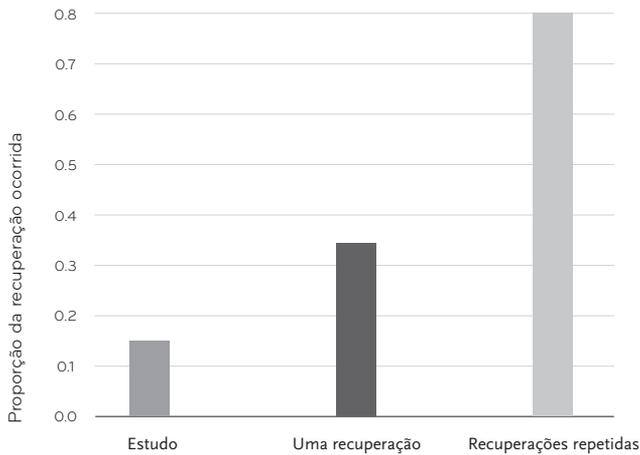
As vantagens da recuperação activa são variadas quando os estudantes recuperam e depois relêem. Em parte, isto acontece porque a tentativa de recuperação melhora a capacidade dos alunos de codificar quando voltam a estudar a matéria, um fenómeno conhecido por efeito «potenciador» da recuperação (ver Grimaldi & Karpicke, no prelo; Karpicke, 2009; Kornell *et al.*, 2009; Wissman, Rawson & Pyc, 2011). Karpicke e Roediger (2010) puseram estudantes a praticar a recuperação de textos educacionais acerca de assuntos científicos. A experiência envolveu algumas condições, mas apenas três são relevantes para o nosso debate. Alunos numa primeira condição leram um texto uma vez apenas num período de estudo, enquanto alunos numa segunda condição leram esse texto, recapitularam o máximo que conseguiram num período de recapitulação e depois releeram brevemente o texto. Numa terceira condição, os alunos recapitularam o texto repetidamente numa série de oito períodos de estudar/recapitular (quatro períodos de estudo e quatro de recapitulação, no total). Uma semana mais tarde, os alunos recapitularam o material de novo para testarem a retenção a longo prazo.

A Figura 2 mostra a proporção de ideias lembradas uma semana após a aprendizagem original.

Praticar a recuperação uma vez duplicou a retenção a longo prazo em relação a ler o texto uma vez (34 % vs. 15 %), e a prática de recuperação repetida aumentou a retenção até 80 % uma semana depois do estudo.

A prática de recuperação activa, conjugada com uma releitura breve entre as tentativas de recapitulação, produziu benefícios a longo prazo naquilo que se quantificou como um período de cerca de meia hora (ver também McDaniel *et al.*, 2009).

Figura 2. Retenção a longo prazo depois de estudar uma vez, de praticar recuperação uma vez (seguida de nova leitura), ou praticar recuperação repetida. Os dados apresentados são adaptados da Experiência 2 de Karpicke e Roediger (2010). A prática de recuperação uma vez duplicou a retenção a longo prazo e a recuperação repetida produziu uma melhoria de 400 % na retenção em relação ao estudo (leitura) único.



A recuperação activa promove a aprendizagem significativa

Um desafio recorrente é estabelecer a eficácia das actividades de aprendizagem baseadas na recuperação com materiais educacionais e testes que reflectam aprendizagens complexas e significativas. «Aprendizagem significativa» é muitas vezes definida por oposição a «aprendizagem rotineira» (Mayer, 2008). Enquanto a aprendizagem rotineira é considerada frágil e transitória, a aprendizagem significativa é forte e contínua. Pensa-se que a aprendizagem rotineira produz conhecimento pobremente organizado, faltando-lhe coerência e integração,

reflectindo-se este facto em falhas na capacidade de efectuar inferências e transferir conhecimentos para novos problemas. A aprendizagem significativa, por outro lado, é vista como produtora de modelos mentais organizados, coerentes e integrados, que permitem às pessoas fazer inferências e aplicar conhecimento adquirido.

É importante lembrar que, em todas as circunstâncias, as pessoas transferem experiências passadas para atender às necessidades do presente. Isto envolve sempre a reconstrução de um conhecimento baseado em pistas disponíveis num contexto específico de recuperação. Resultados identificados como aprendizagens «rotineiras» ou «significativas» podem não reflectir diferenças no que os alunos codificaram, armazenaram ou construíram. Por outro lado, a distinção entre aprendizagem rotineira e significativa articula-se com a semelhança entre cenários de recuperação no presente e experiências de aprendizagem do passado. A capacidade de utilizar conhecimento no presente depende do valor de diagnóstico de pistas de recuperação, quer o objectivo da recuperação seja lembrar um facto, fazer uma inferência ou ainda resolver um problema novo.

Os investigadores seguiram duas abordagens gerais para examinar os efeitos da recuperação activa em avaliações significativas. Uma abordagem consistiu em utilizar questões de avaliação final que diferem das questões empregadas durante a aprendizagem original (por exemplo, Butler, 2010; Chan, 2009; Chan *et al.*, 2006; Hinze & Wiley, 2011; Johnson & Mayer, 2009; Rohrer *et al.*, 2010)². Uma segunda abordagem foi utilizada para delinear avaliações que medissem as capacidades dos

² Existem provas que recuperar uma porção de algum material poderá prejudicar a recuperação de partes não praticadas da matéria (Storm, 2011), mas com materiais educacionais significativos e integrados, a recuperação de parte das matérias aumenta tipicamente a aprendizagem de partes não praticadas das matérias (Chan, 2009; Chan *et al.*, 2006).

alunos para fazer inferências, aplicar conhecimento, resolver novos problemas, ou, por outro lado, medir a coerência e integração dos modelos mentais dos alunos. Já vimos que a recuperação activa promove a aprendizagem de materiais educacionais significativos e que os efeitos são de longo e não de curto prazo. Pesquisas recentes mostraram que praticar a recuperação activa melhora o desempenho medido em avaliações significativas da aprendizagem e que os efeitos podem ser mais vastos do que os produzidos por outras estratégias de aprendizagem.

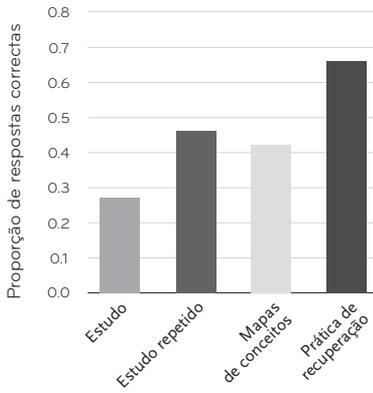
Um estudo recente de Karpicke e Blunt (2011a) examinou os efeitos da recuperação activa na medida de aprendizagem significativa e, mais importante, comparou os efeitos com aqueles produzidos por uma estratégia de aprendizagem popular conhecida como mapeamento de conceitos (Novak & Gowin, 1984). O mapeamento de conceitos implica que os alunos criem diagramas nos quais os conceitos estão representados como nós e as relações entre os conceitos como ligações entre os nós. A actividade solicita um estudo elaborado pois ela é pensada para ajudar os alunos a organizarem e codificarem relações significativas entre conceitos. O mapeamento de conceitos também pode ser utilizado como uma ferramenta para avaliar o conhecimento dos alunos, como se verá a seguir. O mapeamento de conceitos é popular, sendo defendido por muitos educadores. De facto, esta actividade pareceria ser uma ferramenta eficaz para promover o processamento elaborado, mas não tem havido um leque de experiências aleatórias e controladas que examinem as formas mais eficazes de utilizar o mapeamento de conceitos como uma actividade de aprendizagem (ver Karpicke & Blunt, 2011b).

Karpicke e Blunt (2011a) convidaram os alunos a ler textos científicos e a criar um mapa de conceitos ou a praticar a recuperação activa das ideias dos textos. Nas duas situações controladas os alunos leram a matéria uma vez ou repetidamente. Uma semana após a fase de

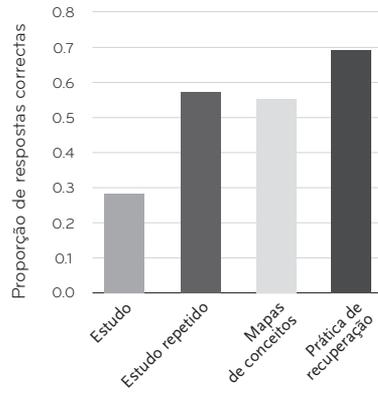
aprendizagem, os alunos responderam a dois tipos de questões que tinham a intenção de avaliar a aprendizagem conceptual significativa: *questões literais* que avaliavam conhecimento conceptual directamente incluído no texto, e *questões inferenciais* que requeriam que os alunos estabelecessem ligações entre conceitos. Como mostram as Figuras 3A e 3B, a prática de recuperação produziu o melhor desempenho em ambos os tipos de questões conceptuais, até melhor do que o estudo elaborado com mapas conceptuais. No entanto, quando os alunos faziam previsões acerca da sua aprendizagem a longo prazo durante a fase inicial de aprendizagem (Figura 3C), eles indicavam que a releitura e o mapeamento de conceitos produziram mais aprendizagem do que a recuperação activa, embora a verdadeira relação fosse a oposta.

Figura 3. Proporção de respostas correctas no final de questões literais de resposta curta (A) e questões inferenciais (B) uma semana depois da aprendizagem, e previsões metacognitivas da aprendizagem feitas durante a fase inicial de aprendizagem (C). A prática de recuperação melhorou a aprendizagem a longo prazo em relação ao estudo elaborado com mapas de conceitos, embora os alunos não estivessem cientes das suas vantagens. Dados de Karpicke e Blunt (2011a) de "Retrieval Practice Produces More Learning Than Elaborative Studying With Concept Mapping,"

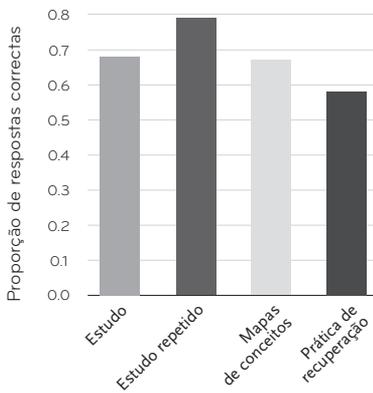
A – Questões literais



B – Questões inferenciais



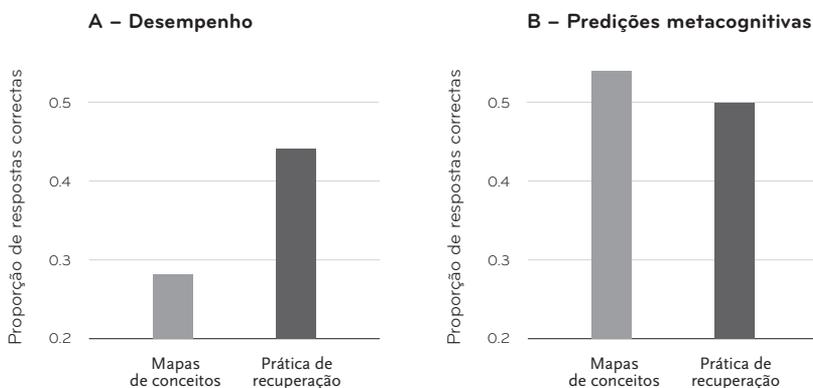
C – Predições metacognitivas



Numa segunda experiência, Karpicke e Blunt puseram de novo os alunos a criar mapas conceptuais ou a praticar recuperação activa enquanto estudavam. Desta vez, no entanto, a avaliação criterial da aprendizagem a longo prazo consistia em criar um mapa conceptual que fosse ele próprio um método de avaliar a coerência e a integração do

conhecimento conceptual dos alunos. Mesmo nesta avaliação do mapa conceptual final (Figura 4A), os estudantes tiveram melhor desempenho quando praticavam a recuperação activa durante a aprendizagem, um resultado que indica que a recuperação activa melhora a aprendizagem conceptual profunda da matéria. No entanto, mais uma vez, os alunos acreditavam, de uma forma geral, que se sairiam melhor após um estudo elaborado do que se praticassem a recuperação, como mostra a Figura 4B. A maior parte dos alunos não esperava que a recuperação activa produzisse mais aprendizagem que o estudo elaborado com mapeamento de conceitos, mas, na realidade, foi isso que aconteceu.

Figura 4. Proporção de respostas correctas num teste final de mapas conceptuais (A) e predições metacognitivas da aprendizagem durante a fase inicial de aprendizagem (B). Os dados apresentados são adaptados da Experiência 2 de Karpicke e Blunt (2011b). Mesmo quando o teste criterial envolveu criar um mapa conceptual, a prática de recuperação durante a aprendizagem original melhorou o desempenho em relação ao estudo elaborado com mapas conceptuais, embora os alunos acreditassem que o mapa conceptual produzisse melhor aprendizagem.



Aprendizagem baseada na recuperação: seguimento

A recuperação é o processo-chave para compreender e promover a aprendizagem. É essencial para compreender a aprendizagem porque todas as expressões do conhecimento envolvem pistas de recuperação que estão disponíveis num determinado contexto. O valor de diagnóstico de pistas de recuperação – o grau no qual as pistas ajudam as pessoas a recuperar um conhecimento específico com a exclusão de outros candidatos competidores – é o factor crítico para toda a aprendizagem. A recuperação é a chave para a promoção de aprendizagem, e a recuperação activa tem efeitos poderosos na aprendizagem a longo prazo. Cada acto de recuperação altera o valor de diagnóstico de pistas de recuperação e melhora a capacidade de cada um de, no futuro, recuperar conhecimento. A recuperação pode melhorar a aprendizagem porque promove uma correspondência entre uma pista e um determinado conhecimento desejado, ou pode melhorar a aprendizagem ao restringir o espectro de procura, o conjunto de candidatos potencialmente recuperáveis que vem à mente no contexto de uma pista (Karpicke & Blunt, 2011a; Karpicke & Zaromb, 2010). A prática da recuperação também mostrou melhorar o processamento organizacional (Congleton & Rajaram, no pelo; Zaromb & Roediger, 2010), sendo este processamento de igual modo necessário para apoiar desempenhos de aprendizagem significativa. Assim, existe um número de potenciais mecanismos através dos quais a recuperação activa pode melhorar a aprendizagem a longo prazo.

A aprendizagem baseada na recuperação é uma perspectiva vasta e geral da aprendizagem. Potencialmente, existem muitas actividades de aprendizagem que podem incorporar a recuperação activa e muitas formas diferentes nas quais a recuperação pode ser integrada nessas

atividades. Por exemplo, discussões de grupo, ensino recíproco e técnicas de questionamento (formais, como fornecer questionários na sala de aula, e informais, como integrar perguntas em palestras) podem utilizar processos de recuperação até um certo ponto. Gastar tempo activamente, tentando recuperar ou reconstruir o conhecimento de alguém é uma forma simples mas poderosa de melhorar a aprendizagem significativa de longo prazo. O desafio central para o futuro da investigação nesta área será continuar a identificar as formas mais eficazes de usar a recuperação como uma ferramenta para melhorar a aprendizagem significativa.

Agradecimentos e nota final

Este trabalho foi apoiado por fundos da National Science Foundation (DUE-0941170) e do Institute of Education Sciences no U.S. Department of Education (R305A110903). As opiniões expressas são as do autor, pelo que não representam pontos de vista do Institute of Education ou do U.S. Department of Education. Agradeço a Stephanie Karpicke a ajuda na preparação do manuscrito.

Referências

- Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, England: Cambridge University Press.
- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1118-1133.
- Chan, J. C. K. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language*, 61, 153-170.

- Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval-induced facilitation: Initially non-tested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, 135, 553-571.
- Congleton, A., & Rajaram, S. (no prelo). The origin of the interaction between learning method and delay in the testing effect: The roles of processing and conceptual retrieval organization. *Memory & Cognition*.
- Grimaldi, P. J., & Karpicke, J. D. (no prelo). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition*.
- Hinze, S. R., & Wiley, J. (2011). Testing the limits of testing effects using completion tests. *Memory*, 19, 290-304.
- Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology*, 101, 621-629.
- Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, 138, 469-486.
- Karpicke, J. D., & Blunt, J. R. (2011a). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331, 772-775.
- Karpicke, J. D., & Blunt, J. R. (2011b). Response to comment on "Retrieval practice produces more learning than elaborative studying with concept mapping". *Science*, 334, 453.
- Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory*, 17, 471-
- Karpicke, J. D., & Grimaldi, P. J. (2011). *Retrieval-based learning: A perspective for enhancing meaningful learning*. Unpublished manuscript, Purdue University.
- Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to longterm retention. *Journal of Memory and Language*, 57, 151-162.
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319, 966-968.
- Karpicke, J. D., & Roediger, H. L. (2010). Is expanding retrieval a superior method for learning text materials? *Memory & Cognition*, 38, 116-124.
- Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language*, 62, 227-239.

- Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 989-998.
- Mayer, R. E. (2008). *Learning and Instruction (2nd Edition)*. Upper Saddle River, NJ: Pearson.
- McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The read-recite-review study strategy: Effective and portable. *Psychological Science*, 20, 516-522.
- Moscovitch, M. (2007). Memory: Why the engram is elusive. In H. L. Roediger, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of Memory: Concepts* (pp. 17-21). New York: Oxford University Press.
- National Research Council (2000). *How People Learn: Brain, Mind, Experience, and School*. J. D. Bransford, A. L. Brown, & R. R. Cocking, (Eds.). Washington DC: National Academies Press.
- National Research Council (2005a). *How Students Learn: Science in the Classroom*. M. S. Donovan, & J. D. Bransford (Eds.). Washington DC: National Academies Press.
- National Research Council (2005b). *Taking Science to School: Learning and Teaching Science in Grades K-8*. R. A. Duschl, H. A. Schweingruber, & A. W. Shouse (Eds.). Washington DC: National Academies Press.
- Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, 10, 389-395.
- Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.
- Novak, J. D. & Gowin, D. B. (1984). *Learning How to Learn*. Cambridge and NY: Cambridge University Press.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, 88, 93-134.
- Roediger, H. L. (1980). Memory metaphors in cognitive psychology. *Memory & Cognition*, 8, 231-246.
- Roediger, H. L. (2000). Why retrieval is the key process in understanding human memory. In E. Tulving (Ed.), *Memory, Consciousness, and the Brain: The Tallinn Conference* (pp. 52-75). Philadelphia: Psychology Press.
- Roediger, H. L., & Guynn, M. J. (1996). Retrieval processes. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 197-236). San Diego: Academic Press.
- Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 233-239.

- Roediger, H. L., & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1, 181-210.
- Roediger, H. L., & Karpicke, J. D. (2006b). Test enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249-255.
- Storm, B. C. (2011). The benefit of forgetting in thinking and remembering. *Current Directions in Psychological Science*, 20, 291-295.
- Tulving, E. (1991). Interview. *Journal of Cognitive Neuroscience*, 3, 89-94.
- Tulving, E., & Pearlstone, Z. (1966). Availability vs. accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 5, 381-391.
- Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review*, 18, 1140-1147.
- Zaromb, F. M., & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, 38, 995-1008.

Leituras recomendadas

- Karpicke, J. D., & Grimaldi, P. J. (2011) (ver Referências). Um resumo recente sobre a aprendizagem baseada na recuperação que abarca tópicos adicionais, incluindo os aspectos metacognitivos da prática da recuperação e os efeitos potenciadores da recuperação na codificação subsequente.
- Moscovitch, M. (2007) (ver Referências). Uma crítica acessível da metáfora da memória como armazém que enfatiza o papel essencial da recuperação na compreensão da aprendizagem.
- Roediger, H. L. (2000). (ver Referências). Uma discussão aprofundada sobre porque a recuperação é o processo-chave para compreender a aprendizagem.
- Roediger, H. L., & Karpicke, J. D. (2006a) (ver Referências). Um resumo histórico e abrangente da investigação sobre os efeitos da recuperação activa na aprendizagem.
- Tulving, E. (1991) (ver Referências). Uma leitura essencial para qualquer estudante da aprendizagem e memória. Uma entrevista na qual Tulving explica por que razão os processos de recuperação devem ser considerados na análise da aprendizagem.

Jeffrey Karpicke é professor associado de Psicologia na Universidade de Purdue e especialista na aplicação da psicologia cognitiva à educação. É mestre e doutorado em Psicologia pela Universidade de Washington em St. Louis.

A sua investigação faz-se na interface entre a ciência cognitiva e a educação, com especial ênfase na importância dos processos de recordação na aprendizagem. O seu objectivo é identificar estratégias eficazes que promovam a aprendizagem significativa a longo prazo e a compreensão.

É autor de mais de vinte publicações, incluindo dois artigos na *Science Magazine*. Em 2011, foi nomeado «estrela em ascensão» pela Associação de Ciências Psicológicas e em 2012 recebeu o prémio *Carreira* da Fundação Nacional para a Ciência. É director do Laboratório de Cognição e Aprendizagem na Universidade de Purdue.

Contactos: Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907-2081; email: karpicke@purdue.edu

QUESTÕES-CHAVE
DA EDUCAÇÃO

EXAMES NACIONAIS: INSTRUMENTOS DE REGULAÇÃO DE BOAS PRÁTICAS DE ENSINO E DE APRENDIZAGEM?

Hélder de Sousa

EXAMES NACIONAIS: INSTRUMENTOS DE REGULAÇÃO DE BOAS PRÁTICAS DE ENSINO E DE APRENDIZAGEM?

Hélder de Sousa³

Os exames constituem um tema de amplo debate que perpassa o meio académico e escolar e assume igualmente grande expressão social, num vasto conjunto de países. Portugal não constitui exceção, sendo recorrente a mediatização do tema, em particular quando são publicitados os resultados de cada época de exames ou, mais recentemente, sempre que são divulgados relatórios nacionais que disponibilizam uma análise global dos resultados, realçando os aspetos mais positivos e mais negativos do desempenho dos alunos.

Partindo de uma breve contextualização do modelo de avaliação externa da aprendizagem em vigor em Portugal, procura-se, através de exemplos de leituras de resultados, em articulação com os resultados de estudos recentes sobre o sistema educativo nacional, contribuir

³ As considerações e opiniões apresentadas nesta comunicação expressam a posição do autor e não devem ser consideradas como posição oficial do Gabinete de Avaliação Educacional (GAVE). O autor escreve segundo o novo acordo ortográfico.

para a compreensão do papel que as provas de avaliação externa podem ter na definição das estratégias dos professores e alunos para, respectivamente, ensinar e aprender, bem como as suas possíveis implicações na qualidade da aprendizagem realizada pelos alunos.

No âmbito desta comunicação, entende-se por «exame» o conjunto de provas de avaliação externa de larga escala, provas essas padronizadas, desenvolvidas e aplicadas por entidades ao nível central da administração educativa (externas às escolas), que obedecem a procedimentos comuns de aplicação e classificação (sendo esta realizada em contexto de total anonimato) e que têm implicações na progressão dos alunos, na certificação e/ou na sua seleção para acesso a outros níveis de ensino. As provas com estas características correspondem ao modelo adotado em Portugal nas atuais provas finais (6.º e 9.º anos de escolaridade)⁴ e nos exames finais, aplicados nos anos terminais do ensino secundário (ES) (11.º e 12.º anos de escolaridade), consideradas provas *high stakes*.

Ainda assim, sempre que relevante, são também considerados os contributos dos resultados das provas de aferição e dos testes intermédios para um entendimento alargado das suas possíveis implicações na regulação das práticas de ensino e de aprendizagem.

1. Exames – um tema sempre controverso

Os exames, em contexto educacional, há muito que constituem motivo de controvérsia académica e social: a discussão em torno do tema é intemporal, inacabada e, porventura, inconclusiva.

⁴ A aplicar também no 4.º ano de escolaridade (1.º ciclo do ensino básico), a partir de 2013.

A polémica entre pais e professores remonta aos primórdios dos exames no nosso país. Segundo Nóvoa (2005, p. 55), «Queirós Veloso explica, em 1898, que a maior facilidade nos exames “tem sido sempre o grande desiderato dos pais, neste país, onde raro se aprende para saber, mas unicamente para passar”». Com contornos similares, a discussão permanece nos nossos dias.

A controvérsia recorrente em torno dos exames, alimentada pelo confronto, mais ou menos extremado, entre os que defendem a importância inquestionável dos exames e aqueles que neles veem uma das principais causas dos insucessos que marcam alguns sistemas educativos, é hoje palco de uma intensa mediatização, implicando vários atores cujos interesses (pelos menos os imediatos), nem sempre coincidentes, são postos em causa pelos resultados globais de um exame específico ou pelo conjunto dos resultados em determinado ano letivo.

Para os professores em particular, e no meio académico em geral, a discussão estrutura-se em torno de posições que enfatizam as vantagens ou as desvantagens que os exames trazem para a concretização de uma melhoria sustentada do desempenho dos alunos e, conseqüentemente, para o aumento da eficácia e da eficiência dos sistemas educativos (Lambert e Lines, 2000; Kellaghan e Madaus, 2003).

Se, por um lado, é aceite que, em geral, os exames contribuem para o estabelecimento de referentes que balizam, a nível nacional, a aplicação do currículo prescrito, na sua extensão, no nível de aprofundamento e, implicitamente, no nível de exigência da aprendizagem que se espera que os alunos realizem, por outro lado, o carácter redutor dos exames escritos, de duração limitada, sustenta críticas relacionadas com o estreitamento do currículo (EURYDICE, 2009).

Naquela linha de argumentação, salientam-se os possíveis impactos na prática letiva e nas estratégias de aprendizagem: ensino

superficial, orientado para o chamado *training to the test*, que absorve de forma excessiva tempo de trabalho com os alunos e sustenta o desenvolvimento de capacidades elementares, orientadas para a reprodução de conhecimento, centradas na memorização de curto prazo.

Ainda no mesmo sentido, sustenta-se uma possível inflação das classificações internas, quando estas concorrem, em conjunto com as classificações dos exames, para a obtenção de uma classificação final. Acresce uma tendência para o imobilismo nas práticas de sala de aula, que concorre para uma dimensão redutora do papel dos exames nos sistemas educativos.

Quando os resultados dos exames são associados a uma avaliação do desempenho profissional ou à avaliação externa das escolas, as suas implicações ganham outra dimensão e uma acrescida visibilidade social (Morton, 2010).

Um dos aspetos centrais da discussão do tema é o contributo dos exames para a elevação da qualidade da aprendizagem, sendo também essa uma das razões que, em alguns países, têm sustentado a sua introdução ou o alargamento da sua aplicação.

Alguma investigação mais recente enfatiza o papel positivo da realização de testes (exames) como ferramenta de reativação da memória. Segundo Salmerón (2011), mais do que a codificação da informação durante a fase de estudo, é o ato de realização do teste que aumenta a eficácia dos processos cognitivos que, no futuro, permitem uma melhor utilização da informação armazenada, bem como uma mais fácil conexão de outra informação com ela relacionada.

Os efeitos da introdução de exames como estímulo externo para a elevação da qualidade dos padrões de ensino e de aprendizagem, para a elevação dos níveis de exigência na formação de professores e, em geral, para a melhoria do funcionamento das escolas e para um maior

reconhecimento social do papel da educação no mundo do trabalho são apontados por Bishop (2001).

Relativizando posições incontestavelmente favoráveis ao papel positivo dos exames, Mons (2009) refere que a investigação desenvolvida internacionalmente não aponta para posições consensuais em torno dos efeitos que avaliações padronizadas possam ter numa efetiva melhoria da aprendizagem. Parece mais seguro aceitar que medidas de política educativa, aplicadas territorialmente de forma diferenciada, estarão na origem do sentido divergente das observações empíricas. As condições de aplicação de uma avaliação padronizada, bem como o modo como dela se apropriam os diferentes atores, poderão ser variáveis mais determinantes para os resultados obtidos do que o processo de avaliação em si mesmo.

Questões como a divulgação e a qualidade dos resultados (nível de desagregação), a forma como são utilizados, enquanto ferramenta de *feedback* positivo para professores e alunos, contribuindo para transformar o desafio de preparação para um exame numa estratégia de aprendizagem, são, aparentemente, pormenores que podem ser decisivos para a obtenção dos efeitos esperados nos países que apostam na realização de exames.

Assim, admitindo como ponto de partida que os exames tenderão a generalizar-se nos vários sistemas educativos (Butt, 2010), será necessário ter sempre presente em que medida as características técnicas dos exames podem interferir nos resultados e condicionar a sua análise e posterior utilização e de que forma a aplicação dos exames se reflete e refletirá na vida escolar, quer dizer, na atividade docente, nas estratégias de aprendizagem usadas pelos alunos e na organização das escolas, na cultura de avaliação socialmente aceite.

2. A difusão dos exames no contexto europeu e em Portugal

A aplicação de exames no ensino difundiu-se largamente ao longo do século passado, sendo hoje uma realidade em numerosos países. No contexto europeu, o recurso aos exames conheceu crescente difusão nos últimos 25 anos. Atualmente, a larga maioria dos países europeus recorre a exames como forma de obter informação sobre o desempenho escolar dos alunos, a fim de monitorizar e avaliar escolas ou ainda avaliar o sistema educativo.

Embora os contextos de aplicação e de utilização dos resultados sejam muito diversos, podem identificar-se aspetos comuns: a larga maioria dos países aplica exames no final do ES (*upper secondary*), predominantemente como medida certificadora de um percurso académico não superior, ou também como marco na transição para o ensino superior⁵. Para este efeito, alguns países, tal como em Portugal, utilizam os resultados dos exames na seleção para o acesso ao ensino superior, com particular impacto no acesso aos cursos e instituições para os quais a procura excede a oferta (EURYDICE, 2012).

Em muitos outros casos os exames são também a solução encontrada para a tomada de decisão sobre percurso escolar dos alunos no final da escolaridade pré-secundária (*lower secondary*, em muitos países)⁶. Portugal, com a transformação das provas de aferição de Língua

⁵ Em dois terços dos países europeus aplicam-se exames no final do ensino secundário com implicações no percurso escolar dos alunos.

⁶ Apenas em cinco países/regiões na Europa, de um conjunto de 35, não se aplicam exames standardizados no final equivalente ao 3.º ciclo do EB em Portugal. Em 17 casos os exames têm implicações no percurso escolar dos alunos. Também são 17 os países/regiões em que os exames servem múltiplas funções, como seja a supervisão do desempenho de escolas ou do sistema educativo ou ainda a deteção de dificuldades de aprendizagem dos alunos.

Portuguesa e de Matemática do 3.º ciclo do ensino básico (EB) em exames nacionais, em 2005, passou a estar em linha com aquela tendência europeia (EURYDICE, 2009).

Atualmente, o contexto de crescente autonomização das escolas, que pode implicar uma maior heterogeneidade dos critérios e dos procedimentos relativos à avaliação interna dos seus alunos, tem igualmente suscitado o crescente interesse pelo estabelecimento de sistemas de avaliação externa centrados em exames (EURYDICE, 2009). Pretende-se instituir um sistema de avaliação independente de cada escola, que assegure a manutenção de padrões de referência à escala nacional e um mínimo de homogeneidade na gestão do currículo, podendo-se assim garantir comparações do desempenho entre escolas e/ou regiões.

A avaliação externa também é usada, de forma generalizada, como um instrumento para monitorizar a eficiência e eficácia dos sistemas educativos. A pressão que ao nível do universo das entidades empregadoras é implícita ou explicitamente exercida no sentido de dispor de uma informação complementar à fornecida pelas escolas, assente em critérios normalizados e uniformes à escala de cada país, constitui, em algumas sociedades, um fator acrescido para o reforço da relevância dos exames.

Também a ocorrência de resultados menos positivos em estudos de avaliação internacional de larga escala, como o PISA, o TIMSS ou o PIRLS, em alguns países onde a avaliação externa não tinha tradicionalmente grande expressão, contribuiu para a introdução ou para uma mais expressiva presença de exames externos (EURYDICE, 2009).

A adoção de exames nos vários países pode ser caracterizada pela combinação de múltiplas opções, de que se destacam: anos de escolaridade/ciclos em que são aplicados; leque de disciplinas sujeitas a exame; número de exames obrigatórios ao longo do percurso escolar dos alunos; peso relativo das classificações internas e externas na

determinação das condições de classificação, certificação e/ou de seleção; utilização dos resultados dos exames.

No contexto europeu, Portugal apresenta um modelo de avaliação externa dos alunos que revela grande equilíbrio, atendendo à combinação das componentes atrás referidas. Apesar de existirem exames desde o final do século XIX (Martins, 2008), a sua aplicação de forma mais rigorosa e sistemática, em moldes mais semelhantes aos atuais, surgiu na década de 1940, a par da Islândia, do Reino Unido (Escócia e Irlanda do Norte), do Luxemburgo e da Holanda (EURYDICE, 2009).

Os exames, suprimidos temporariamente após 1974, são retomados duas décadas depois no final do ES (12.º ano) e, em 2005, como atrás referido, passam a aplicar-se também no final da escolaridade obrigatória (9.º ano, à data). Os exames aplicados no EB constituem uma herança de provas de aferição (*low stakes*), aplicadas em anos anteriores. Mais recentemente (2012), as provas de aferição deram lugar a provas finais (*high stakes*), no final do 2.º ciclo do EB.

No que se refere ao leque de disciplinas sujeitas a avaliação externa, Portugal apresenta uma situação coincidente com a maioria dos demais países europeus. Incidência nas disciplinas de língua materna e de Matemática nos diversos ciclos do ensino básico; no ensino secundário aplica-se a todos os alunos na disciplina de Português, tendo os alunos de realizar ainda mais três provas de exame em disciplinas específicas, de acordo com o seu plano de estudos.

Complementarmente, com exceção do ano transitório da primeira aplicação dos exames nacionais do 3.º ciclo e das provas finais do 2.º ciclo do ensino básico, a classificação interna contribui em 70 % para o cálculo da classificação final de frequência. As classificações dos exames nacionais/provas finais contribuem com os restantes 30 %.

Esta repartição do peso relativo da avaliação baseada na informação de cada professor e da avaliação concebida e aplicada em contexto externo é mais um sinal de equilíbrio no modelo de avaliação e de certificação da aprendizagem. Trata-se da solução adotada aquando da reintrodução dos exames terminais do ES, nos anos 90 do século passado, também utilizada nos restantes ciclos de ensino.

Em suma, pode dizer-se que o sistema de avaliação nacional evidencia princípios que lhe garantem uma elevada qualidade concetual – abrangência, coerência e continuidade –, concretizados em três dimensões essenciais: a *vertical* – avaliação em vários níveis de escolaridade; *horizontal* – articulando avaliação interna e externa, sumativa e formativa, baseada no currículo; *longitudinal* – ao longo do percurso escolar dos alunos (Pellegrino *et al.*, 2001).

3. Os resultados dos exames: potencialidades e alguns constrangimentos na sua leitura e interpretação

Fazer um uso adequado dos resultados dos exames implica, por um lado, conhecer as limitações que decorrem das múltiplas variáveis externas ao próprio exame e que afetam a sua validade e a fiabilidade dos seus resultados (Butt, 2010; Kellaghan e Madaus, 2003; Tindal e Haladyna, 2002). Por outro lado, enquadrando aquelas contingências nos processos de análise efetuados, os resultados podem constituir, se devidamente apropriados e utilizados, indicadores úteis para as tomadas de decisão ao nível de opções pedagógicas e didáticas que promovam uma contínua melhoria da aprendizagem.

A apropriação dos resultados, e a sua utilização para os fins referidos, é tarefa que deve envolver a administração educativa, as organizações escolares, os professores, mas também os alunos e os pais.

A disponibilização de resultados com diferentes níveis de desagregação permite a cada um dos atores um uso diferenciado dos mesmos, no quadro dos diferentes papéis que cada um pode e deve desempenhar no sistema educativo.

Não se pretendendo apresentar uma análise exaustiva dos constrangimentos e das potencialidades na utilização dos resultados, salientam-se os seguintes aspetos, enquadrando o contexto observado em Portugal:

- i. níveis de desagregação dos resultados, política de divulgação e sua utilização;
- ii. comparabilidade dos resultados;
- iii. validade dos exames e fiabilidade dos resultados.

i. Níveis de desagregação dos resultados, política de divulgação e sua utilização

A perceção do contributo da avaliação externa para a melhoria da qualidade formativa dos alunos implica a disponibilidade de informação de elevada qualidade e adequado nível de desagregação – resultados globais nacionais, resultados desagregados por região, escola, turma, item, aluno – bem como a sua divulgação regular e sistemática.

Em Portugal, os resultados dos exames dados a conhecer às escolas, aos professores e divulgados publicamente limitaram-se, durante mais de uma década, às classificações médias nacionais, à distribuição de frequência e um indicador de correlação entre as classificações internas e as dos exames. Esta informação nada diz sobre o desempenho específico dos alunos e não permite uma fina explicitação dos domínios da aprendizagem com fragilidades e com necessidade de especial atenção, através de uma reformulação das estratégias de ensino e de aprendizagem.

No quadro das provas de aferição, aplicadas desde 2001⁷, só a partir de 2007 passaram a ser divulgadas as classificações dos alunos e os resultados globais. Os resultados por item passaram também a estar disponíveis, a nível nacional (relatórios nacionais, públicos) e por turma/escola (relatórios de escola, de acesso restrito).

No caso dos testes intermédios⁸, concebidos externamente mas aplicados e classificados nas escolas, os resultados de cada aluno, por item, ficam de imediato disponíveis para os respetivos professores. Os resultados de cada turma podem ser comparados internamente nas escolas, e os resultados de cada escola podem também ser confrontados com resultados regionais e nacional, através dos relatórios técnicos por prova. Desde 2009 são publicados relatórios nacionais.

No caso dos exames nacionais, foram pela primeira vez divulgados às escolas os resultados dos seus alunos desagregados por item em 2010. Desde 2009, estão também disponíveis ao público relatórios nacionais.

Em 2010, foi melhorada a informação dos resultados das provas de aferição. Além dos resultados globais de cada aluno, passou também a ser divulgada informação desagregada por domínio (Língua Portuguesa) e tema (Matemática), medida que em 2012 foi também adotada nas provas finais do 6.º e do 9.º anos. Outro exemplo é o que consiste na apresentação dos resultados dos testes intermédios de Língua Portuguesa e de Matemática do 2.º ano de escolaridade. A informação disponibilizada aos professores e aos pais inclui, além de um resultado

⁷ Aplicadas no 4.º ano entre 2001 e 2012, no 6.º ano, entre 2001 e 2011 e no 9.º ano em 2002, 2003 e 2004. Em 2007, as provas deixaram de ser realizadas em regime de anonimato, como nos anos anteriores.

⁸ Aplicados desde 2005 na disciplina de Matemática A do ensino secundário, abrangem atualmente 12 disciplinas do 2.º, 8.º, 9.º, 10.º, 11.º e 12.º anos de escolaridade.

global e também por domínio/tema, uma descrição sumária do nível de consecução do objetivo de cada item, por aluno.

A divulgação de informação, mais pormenorizada e complementar do que o habitual conhecimento de um resultado global, o qual apenas indica a posição relativa do aluno por referência às normas de classificação estabelecidas para o seu nível de ensino, apresentam significativas vantagens: a possibilidade de um diagnóstico mais preciso das dificuldades de aprendizagem de cada aluno, o que pode constituir um estímulo para um envolvimento informado dos professores e dos pais na definição de estratégias de ação que concorram para a superação das dificuldades detetadas.

A divulgação da informação com o nível de desagregação referido poderá sustentar intervenções de apoio educativo, em contextos formais ou informais, dirigidas a áreas específicas da aprendizagem, com ganhos de eficácia e de eficiência em relação ao tempo utilizado, por todos os atores. Poderá ainda evitar-se a tradicional e fácil tendência para rotular alunos como sendo «fracos» nesta ou naquela disciplina. Havendo casos em que tal se pode aceitar, há muitos outros em que as dificuldades se localizam em aspetos concretos, em conteúdos ou em capacidades específicas, podendo nestes casos evitar-se a inclusão dos alunos em programas de apoio indiferenciados e de duração muito superior à necessária.

Globalmente, o grau de desagregação da informação hoje disponível e os destinatários a quem se dirige, envolvendo de forma muito direta escolas, professores e pais, coloca Portugal como um dos países europeus que mais evoluíram neste domínio específico (Santiago *et al.*, 2012).

Todavia, a política de divulgação de resultados atrás descrita é, como se percebe, muito recente. A cultura de apropriação dos resultados pelas organizações escolares, pelos professores e também pelos

país é, ainda, incipiente. Se, por um lado, há uma clara evidência de uma crescente preocupação das escolas na recolha sistemática de dados (internos e externos), por outro lado, a informação recolhida é objeto de análises, em regra, ainda muito superficiais (Santiago *et al.*, 2012).

ii. Comparabilidade dos resultados – diferentes dimensões

A comparabilidade dos resultados pode assumir múltiplas dimensões, sendo hoje um dos temas centrais do interesse que a comunidade, em geral, tem relativamente ao conhecimento dos resultados da avaliação externa.

Através do estudo dos resultados podemos comparar o desempenho de escolas ou de regiões, permitindo que cada entidade envolvida se situe em relação a uma medida padronizada de âmbito nacional. Podem-se comparar escolas e/ou regiões em função dos resultados obtidos numa mesma prova (ou conjunto de provas) realizada num dado momento, ou longitudinalmente, considerando uma série de resultados alcançados em diferentes anos; ou ainda, em processo de avaliação interna (formal ou informal), pode ser comparada a evolução dos resultados que os alunos de uma mesma escola alcançam em anos sucessivos.

A leitura dos resultados nos contextos atrás referidos não deve, no entanto, ignorar os efeitos perversos gerados por comparações cegas, sem contextualizar cada escola ou região de um ponto de vista social e económico. Ou ainda sem ter em conta, por exemplo, a capacidade de cada escola em preservar, no final de um ciclo de escolaridade, em sede de prova final/exame nacional, a coorte de alunos que o iniciou.

Os conhecidos *rankings* de escolas, tão do agrado de alguma comunicação social e de algumas escolas, em regra com públicos predominantemente provenientes de contextos mais favorecidos, ou que

promovem implicitamente mecanismos de seleção dos alunos internos admitidos a exame, são objeto de forte crítica em vários países e mesmo proibidos noutros (Butt, 2010; EURYDICE, 2009; Lambert, 2000).

O conceito de valor acrescentado em educação (Butt, 2010; Lambert, 2000) constitui uma variável essencial para uma correta utilização dos resultados da avaliação no âmbito referido. De notar ainda que uma correta comparação dos resultados de avaliações externas de carácter público, sujeitos a variações decorrentes da natureza sempre original das provas e ainda das variações relevantes observadas no desempenho dos alunos (quando existem), deve privilegiar o uso de valores índice, tendo por referência as médias nacionais de cada ano, e não valores absolutos dos resultados médios publicados.

iii. Validade dos exames e fiabilidade dos resultados

O carácter público das provas, como sucede em Portugal, é precisamente um dos aspetos mais sensíveis no processo de apropriação e uso dos resultados, principalmente quando através de comparações longitudinais se pretende fazer um retrato da evolução de um dado sistema educativo.

Sendo as provas públicas sempre originais, não sendo possível recuperar itens anteriormente publicados, é preciso ter presente que a fiabilidade dos resultados depende, em última análise, da estabilidade concetual dos enunciados e critérios de classificação utilizados.

Ainda que a estabilidade dos resultados das provas, em termos longitudinais, possa constituir um valor inerente à sua validade, é preciso ter presente que a valorização dessa estabilidade, em detrimento da importância que deve ser dada às características técnicas das provas, assenta na premissa de que o desempenho das sucessivas coortes de

alunos se deverá manter inalterável. Este facto contraria o principal objetivo de qualquer sistema educativo, que se entende ser o da progressiva elevação da qualidade da aprendizagem realizada pelos alunos.

Se aquela circunstância for verdadeira, ocorrerá uma pequena variabilidade interanual dos resultados. Porém, se se registar uma melhoria gradual e sustentada do desempenho dos alunos (ou uma tendência de sentido inverso), só uma elevação (ou redução) dos resultados, tendo por base provas concetualmente estáveis, permite aferir o sentido e a intensidade das alterações do nível de desempenho que se pretende observar.

Uma variação dos resultados globais mais acentuada do que o habitual, em dois ou mais anos consecutivos, não deverá implicar, obrigatoriamente, uma quebra de confiança nos exames: a referência para garantir a confiabilidade dos resultados deve assentar na comparação dos enunciados e dos critérios de classificação das provas e não numa imutável persistência de resultados similares, em anos consecutivos.

Ou seja, não sendo a conceção de exames uma ciência exata, a sua validade depende também da forma como são lidos os resultados. Segundo Stobart (2008, p. 174), “um teste bem construído torna-se inválido se os resultados são mal entendidos ou interpretados de forma errada”. Muito do debate que anualmente se gera em torno dos resultados dos exames, em Portugal, alimentado por opiniões mal informadas ou por leituras, intencionalmente ou não, desfocadas, ignora a realidade atrás descrita.

Uma derradeira observação prende-se com a ausência de uma relação explícita e formal entre o currículo e os graus de desempenho esperados nas prestações dos alunos na sua avaliação (interna e externa). Esta circunstância, também observada em Portugal, constitui uma acrescida dificuldade no estabelecimento de um referencial objetivo para a conceção de provas de avaliação externa e para a comparação longitudinal dos seus resultados.

No caso das provas nacionais, os vários constrangimentos que atrás se enunciaram são minimizados pela estabilidade de características essenciais de cada exame, como sejam a extensão, a estrutura, a tipologia dos itens e o peso relativo da distribuição da cotação total pelos tipos de itens, as operações mentais mobilizadas a partir dos verbos de comando utilizados nos itens, as características e a complexidade dos suportes dos itens, elementos-chave para uma comparação das provas que permita aferir a sua validade global.

Ainda no sentido de minimizar o impacto do carácter público das provas na variabilidade interanual dos resultados, a análise psicométrica dos resultados dos itens publicados e a manutenção de critérios estáveis, no que se refere aos parâmetros que interferem no grau de dificuldade dos itens (conteúdo testado; estímulo; tarefa a desenvolver; resposta esperada) (Leong, 2006), constituem medidas complementares que visam assegurar padrões de exigência similares em diferentes anos. Desta forma, torna-se possível considerar o desempenho dos alunos como a variável-chave na explicação da evolução longitudinal dos resultados, aumentando o grau de confiança dos exames como medida da qualidade da aprendizagem.

Um outro aspeto a ter em atenção é o facto de todas as provas aplicadas (e respetivos critérios de classificação) estarem disponíveis na Internet, a que acresce uma enorme profusão de publicações que replicam e complementam esse acervo com itens e propostas de resolução. Sendo muito positivo o facto de toda a informação ser do domínio público, a familiarização dos alunos com os exames pode ser entendida como uma variável residual, não devendo ser totalmente desconsiderada quando se pretende fazer leituras longitudinais de resultados.

O treino para os exames apenas como estratégia de última hora terá benefícios residuais para a qualidade da aprendizagem futura e

poderá contribuir para sustentar uma estagnação ou inflação de resultados que não espelham uma consistente elevação da qualidade dessa aprendizagem. No entanto, se utilizado como meio de autorregulação (treino de rotinas, detecção e superação atempada de dificuldades), seja através do trabalho individual, em pares e através de processos de *feedback* interativos, professor-aluno e aluno-professor, poder-se-á esperar que uma subida dos resultados reflita uma sustentada elevação dos padrões de desempenho dos alunos.

Outros dois aspetos apontados como fragilidades dos exames são os que se prendem, por um lado, com os efeitos de erro que a intervenção humana no processo de classificação pode introduzir nos resultados e, por outro lado, com a qualidade e a objetividade dos critérios de classificação de cada prova. Estas duas variáveis ganham particular expressão nos exames que incluem itens de construção, sendo irrelevantes no que se refere aos itens de seleção, onde se incluem os de escolha múltipla.

A fiabilidade da classificação será sempre condicionada pela variação da consistência dos critérios que cada classificador adota durante o período em que decorre a classificação e pela diversidade de critérios pessoais que cada classificador, intrinsecamente, mobiliza nas suas tomadas de decisão.

No caso nacional, podemos considerar três medidas para minimizar as referidas fragilidades: forte investimento na qualificação e no acompanhamento dos professores classificadores⁹; garantia da reapreciação

⁹ Os classificadores das provas finais do EB estão enquadrados num processo de supervisão da classificação que implica formação específica na tarefa e acompanhamento especializado durante o período em que decorre a classificação. Idêntico processo esteve em vigor em seis disciplinas do ES até 2011. Atualmente os classificadores de todas as provas do ES estão integrados num programa de formação. O processo de

das provas *high stakes* (em média, a percentagem de provas reapreciadas no ES é da ordem dos 2 % a 3 %); melhoria contínua da qualidade dos critérios de classificação (nos últimos vinte anos passou-se de um modelo de critérios de classificação que consistiam numa descrição genérica dos procedimentos a adotar para critérios específicos por item, incluindo descritores associados a níveis de desempenho).

4. O que nos dizem os resultados das provas de avaliação externa realizadas em Portugal, nos últimos cinco anos

Salvaguardadas as limitações inerentes às comparações longitudinais de resultados de provas públicas, atrás enunciadas, e desconsiderados resultados atípicos e pontuais que enviesam a validade e a sua consequente comparabilidade, observados num passado ainda recente¹⁰, ou ainda os efeitos de alterações estruturais ou concetuais de provas, também pontuais, consequências de ajustamentos curriculares

classificação está concebido de molde que todas as respostas possam ser objeto de análise por mais do que um elemento, para reduzir o erro e a consequente quebra de fiabilidade.

¹⁰ Exemplos de resultados considerados atípicos: Matemática A, valores de 2008 (14 valores) e dos dois anos subsequentes (11,7 e 12,2, respetivamente), anos de ajustamento gradual aos valores de referência observados em anos anteriores e retomados em 2011 e 2012; Português, valor de 2011 (9,6 valores). Este valor está associado a um desvio negativo considerável observado em dois itens do Grupo II (Funcionamento da Língua – Gramática) em relação ao histórico de resultados de itens que, em anos anteriores, mobilizaram conhecimentos análogos – em 20 pontos possíveis, a classificação em 2009 foi 14,6, em 2010 12,4, e em 2011 apenas 7,3. Este desvio contribuiu para uma redução da média global em cerca de 5 pontos, em comparação com o valor esperado.

externos ao processo de conceção das provas, observa-se uma relativa estabilidade das classificações médias dos alunos, no universo das disciplinas sujeitas a provas externas¹¹.

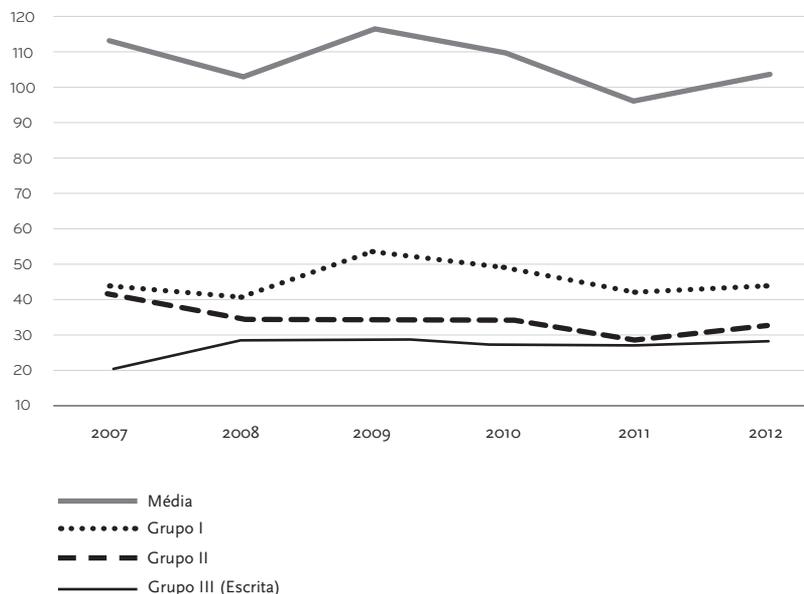
Os resultados dos exames nacionais do ES (alunos internos – 1.^a fase), realizados entre 2007 e 2012, mostram-nos o seguinte:

- a variação média dos resultados de um ano para outro, nas disciplinas consideradas situou-se entre 0,8 valores (2009-2010) e 1,1 valores (2006-2007);
- a amplitude dos resultados por disciplina variou entre os 0,1 valores (Matemática A) e 3,1 valores (Física e Química A), situando-se a maior parte entre 1 e 2 valores;
- a inexistência de tendência sustentada de aumento de resultados em qualquer das disciplinas consideradas.

A evolução dos resultados do grupo III do exame de Português do 12.^o ano, apresentada na figura 1, constitui um dos muitos exemplos que ilustram a estabilidade de resultados da avaliação externa, atrás referida.

¹¹ Entende-se como estável (atendendo à natureza pública e à consequente originalidade de todas as provas) uma variação interanual não superior a 1 valor nos resultados da 1.^a fase dos exames do ES com um afluxo médio igual ou superior a 2500 alunos internos: Português, Matemática A, Matemática Aplicada às Ciências Sociais, Física e Química A, Biologia e Geologia, Geografia A, História A, Economia A, Geometria Descritiva A e Desenho A.

Figura 1. Evolução da classificação média global e por grupos (alunos internos, 1.ª fase – pontos entre 0 e 200)



A variação da média global na disciplina, ainda que dentro do intervalo de 1 a 2 valores é essencialmente reflexo da variação interanual dos resultados médios gerados no grupo I da prova. Este grupo, cuja cotação representa 50 % do total da cotação da prova, inclui questões que, embora obedecendo a critérios de conceção estáveis, estão obrigatoriamente associadas à leitura e interpretação de excertos de textos de diferentes obras/autores. A seleção destes constitui o principal fator que condiciona os resultados médios globais.

Já no que se refere aos resultados do grupo III, constituído por um item de escrita cuja formulação e cujos critérios de classificação

se mantêm constantes no intervalo temporal considerado, observa-se uma variação das pontuações médias irrelevante: tendo o item uma cotação de 50 pontos, a pontuação mais alta da série foi 28 pontos e a mais baixa 21 pontos; estes valores representam uma percentagem da cotação média, em relação à classificação total, da ordem dos 55 %.

No âmbito da capacidade da escrita, principalmente na produção de textos de carácter expositivo-argumentativo, nos aspetos associados à estrutura, à coesão e até mesmo ao nível da pontuação e da ortografia, continuam a ser notórias preocupantes dificuldades na produção dos alunos. Os alunos não «descolam» de uma prestação pouco acima de um valor que corresponde a metade da cotação disponível no item, num quadro de provas públicas em que a estrutura do item e as capacidades mobilizadas se têm mantido inalteradas.

As evidências aqui expostas são também extensíveis aos resultados obtidos no exercício de escrita proposto nas provas de aferição de Língua Portuguesa no 4.º e no 6.º anos de escolaridade, e nos exames nacionais desta disciplina no 9.º ano.

De notar que uma das causas explicativas de desempenhos menos sucedidos nos itens de resposta de construção extensa, apontadas nos relatórios nacionais dos exames e dos testes intermédios, prende-se com o fraco domínio da escrita, capacidade transversal cujo desenvolvimento é apontado como essencial para a elevação da qualidade das respostas (GAVE, 2009, 2010 e 2011).

Centrando o foco da leitura de resultados ao nível do item (unidade de análise que deve ser privilegiada em estudos longitudinais em provas públicas), e considerando os resultados nas respostas aos itens que mobilizam capacidades de nível mais elevado, que envolvem operações mentais mais complexas (por exemplo, leitura inferencial, resolução de problemas, delineação de estratégias de resposta que implicam

várias etapas), verifica-se uma situação que não se afigura diferente da exemplificada anteriormente.

Embora de forma não tão sistemática como a observada em relação ao domínio das capacidades de nível mais elevado, o mesmo se verifica em alguns resultados de respostas a itens cuja correta resolução assenta na mobilização de conhecimentos e de capacidades de nível elementar, como, por exemplo, a reprodução de conhecimento factual, o uso adequado de terminologia específica ou a realização de tarefas que assentam em procedimentos que implicam rotinas básicas (GAVE 2009, 2010 e 2011).

Uma análise pormenorizada das fragilidades reveladas pelos resultados da avaliação externa que sustentam o diagnóstico de uma preocupante ausência de sinais de melhoria pode ser vista nas várias edições dos relatórios nacionais das provas de aferição e dos exames nacionais e testes intermédios.

A evidência de que no contexto nacional se observa uma estagnação dos resultados da avaliação externa tem também paralelo na avaliação interna. A média das classificações internas de frequência (CIF) dos alunos do ES, obtida nos últimos três anos nas disciplinas com exame nacional¹², situa-se entre 13 e 15 valores, não se registando variações interanuais por disciplina superiores a um valor (GAVE/JNE).

Não obstante a diferença de valores médios entre os resultados da avaliação interna e da avaliação externa ser variável de disciplina para disciplina, fenómeno que aqui não se pretende abordar, regista-se uma grande estabilidade no intervalo observado entre a média da CIF e a média dos resultados dos exames em cada disciplina, o que é consistente com o quadro de estagnação que atrás se ilustrou.

¹² Consideram-se as disciplinas enumeradas na nota 8.

5. Exames, estratégias de ensino e de aprendizagem – Portugal e o contexto internacional

Responder à interrogação que o título da comunicação apresenta não é tarefa fácil. No imediato, entendendo como boas práticas aquelas que podem gerar uma elevação dos padrões de qualidade da aprendizagem, parece evidente que resposta terá de ser negativa. Mas igualmente negativa será a resposta à pergunta: serão os exames os principais responsáveis pela cristalização dos resultados nacionais?

A investigação mostra-nos uma persistente falta de integração das diferentes modalidades de avaliação vigentes no nosso sistema educativo, um enfoque em modelos tradicionais de ensino, que conduzem a práticas de sala de aula orientadas para a preparação para os testes, e uma avaliação de cariz predominantemente sumativo, que reflete lacunas na formação inicial dos professores, no que se refere a conceções e práticas de avaliação alternativas às tradicionais (Fernandes, 2009; Santiago *et al.*, 2012).

O quadro atrás descrito, próprio de sistemas dominados por avaliações *high stakes*, merece especial discussão. Recordemos que, em Portugal, a avaliação externa *high stakes* foi retomada há cerca de quinze anos apenas no ES, e há sete anos no ano terminal do EB (9.º ano), que abrange duas das disciplinas do currículo do EB, e que, para a conclusão do ES, os alunos só fazem quatro exames. Até 2012, em doze anos de escolaridade, os alunos apenas realizavam exames em três anos (9.º, 11.º e 12.º anos). No entanto, não existe evidência de que as práticas de ensino sejam diferenciadas em função da pressão externa dos exames, pelo menos de forma generalizada.

O modelo cultural de avaliação vigente está fortemente associado aos exames e à ideia de que só se avalia quando se atribui um resultado. É um modelo marcado pelas vivências escolares dos adultos,

professores e pais, impõe-se aos discentes de hoje e sustenta-se, socialmente, pela apetência por resultados, descurando, no essencial, preocupações com a dimensão processual da aprendizagem (Heritage, 2009), reconhecida como relevante para a geração de bons resultados.

Para os pais, em regra, importa saber se o seu filho teve ou não «positiva»; menos importante é saber o porquê da distância em relação a um desempenho de excelência, ou ainda antecipar que um resultado próximo da linha de corte pode ser indicador de problemas de aprendizagem futuros. Entre os alunos, perante qualquer tarefa proposta pelo professor, ainda é muito frequente a pergunta: «Professor, isso conta para nota?» Para a imprensa, não são os casos inovadores, os casos de experiências educativas de sucesso, que também existem, que fazem notícias de primeira página.

Parece evidente que o modelo do *training to the test* se afigura esgotado. Nos países com forte tradição em avaliações na sala de aula, de que o Reino Unido, a Holanda ou os países nórdicos são exemplos, assiste-se a numerosas experiências educativas que corporizam os resultados de investigações que apontam para uma verdadeira integração da avaliação formativa e sumativa (Butt, 2010; Harlen, 2006).

Uma perspetiva integrada, como a referida, pressupõe uma adequada utilização dos resultados da avaliação sumativa (externa) para informar e modelar o ensino e o modo como se aprende. A informação gerada pela avaliação interna, que deve ter cariz essencialmente formativo, pode ser usada para tomadas de decisão nas escolas e no sistema educativo (Butt, 2010; Looney 2011). Embora não gere, de momento, consenso no meio académico, essa abordagem é objeto de ampla investigação aplicada, com resultados muito promissores (Harlen, 2006; Looney, 2011).

O sucesso de qualquer estratégia de sala de aula que vise explorar as potencialidades da perspetiva de integração referida é indissociável da

partilha de *feedback* de qualidade (Lambert, 2000). Em Portugal, a fraca predisposição dos professores para partilhar, com pais e alunos, as decisões e os critérios relativos à avaliação desenvolvida em sala de aula (Fernandes, 2009) limita as oportunidades para um *feedback* capaz de desencadear processos de aprendizagem sustentados pelas informações dessa mesma avaliação. A prática mais comum consiste em centrar a informação devolvida nos resultados quantitativos dos testes e não na dimensão qualitativa da aprendizagem. Os professores devolvem resultados, apresentam as soluções, mas pouco ou nada é feito no sentido de promover um envolvimento dos alunos na apropriação dos seus processos cognitivos, ajudando-os a desenvolver capacidades metacognitivas, essenciais para sustentar aprendizagens futuras mais consistentes (Black e Wiliam, 2009; Heritage, 2009).

Adicionalmente, a aprendizagem autorregulada tem, reconhecida, a maior relevância na promoção de melhores desempenhos, na elevação dos níveis de motivação dos alunos e na sua predisposição para continuar a aprender, um valor da maior importância na consolidação do processo de aprendizagem ao longo da vida (Clark, 2012; Hopfenbeck, 2011). A autoavaliação ou a avaliação por pares, por exemplo, suscitam um maior envolvimento dos alunos nos processos de aprendizagem e de avaliação (Butt, 2010), com impacto no desenvolvimento de capacidades como a organização e o reprocessamento da informação ou a resolução de problemas (Clark, 2012; Pellegrino, 2006).

No contexto europeu, os países que melhor têm sabido aproveitar as sinergias entre as componentes interna e externa da avaliação, centradas, respetivamente, nas suas dimensões formativa e sumativa, são, em regra, os que mostram melhores resultados nos estudos internacionais de avaliação de alunos de larga escala, como o PISA.

Em Portugal, como noutros países, o investimento em recursos financeiros, humanos e materiais em apoio da realização de exames gera a

legítima expectativa de uma progressiva elevação dos padrões de aprendizagem (Pellegrino, 2006). A não concretização desse objetivo não deve, no entanto, constituir pretexto para radicalizar posições contra os exames.

Nem os exames, nem qualquer outra forma de avaliar estão isentos de imperfeições. Elevar a qualidade da aprendizagem, e assim esperar uma melhoria sustentada dos resultados dos exames, parece implicar uma mudança de paradigma no que se refere à forma como socialmente se percebe a educação, a avaliação e a partilha de resultados. No que se refere em particular ao papel dos exames, parece justificar-se uma abordagem alternativa: *working with the tests (and using tests results) to improve learning*.

Conclusão

Os resultados dos exames realizados pelos alunos portugueses nos últimos cinco anos não mostram, em geral, sinais de uma efetiva melhoria da qualidade da aprendizagem, facto consistente com os resultados da avaliação interna no mesmo período. Criticar os exames como responsáveis pela persistência de práticas de sala de aula e por comportamentos dos alunos pouco consistentes com a realização de uma aprendizagem de qualidade é o caminho mais fácil.

Contudo, dificilmente se pode sustentar aquela posição se se atender: ao peso relativo que os resultados dos exames representam na classificação final dos alunos; ao número de anos de escolaridade e de disciplinas em que estes se aplicam; à importância que, formalmente, assume a avaliação interna e formativa na arquitetura do processo de avaliação.

O modelo de avaliação de alunos socialmente consolidado, assente num trabalho de sala de aula excessivamente focado no treino para os exames, e que se replica também nos anos e disciplinas em que não se

aplicam exames, reflete um imobilismo nas concepções que os diferentes atores, professores, pais e alunos, têm em relação aos atos de avaliar ou de ser avaliado.

Num cenário em que a supressão dos exames se afigura implausível, é necessário encontrar nas escolas, nas salas de aula, nas famílias, o tempo e o espaço para uma reflexão sobre o modo como se deve ensinar e aprender, explorando as potencialidades da avaliação interna e externa, formativa e sumativa, onde se incluem os exames, na medida em que qualquer delas pode contribuir, embora de forma diferente, para concretizar uma aprendizagem de maior qualidade, no quadro de uma desejável elevação da qualificação da população.

Melhorar os padrões de qualidade da aprendizagem é uma tarefa complexa e morosa. Mudar o paradigma de avaliação dominante na sociedade, também.

Para já, de forma mais simples, um primeiro passo pode ser dado no sentido de uma progressiva apropriação e utilização da informação gerada pelos resultados da avaliação externa, os quais têm sido amplamente divulgados a todos os interessados.

Referências bibliográficas

- Bishop, J., Mañe, F. & Bishop, M. (2001). «How External Exit Exams Spur Achievement». *Educational Leadership*, vol. 59, n.º 1, pp. 58-63. Consultado em junho 2009 em <http://www.ascd.org/publications/educational-leadership/sept01/vol59/num01/How-External-Exit-Exams-Spur-Achievement.aspx>.
- Black, P., Wiliam, D. (2009). “Developing the Theory of Formative Assessment”, *Educational Assessment Evaluation and Accountability*, Vol. 21, n.º 1, pp. 5-21.
- Butt, G. (2010). *Making Assessment Matter*. Londres: Continuum Int. Publishing Group.
- Clark, I. (2012). «Formative Assessment: Assessment Is for Self-regulated Learning». *Educational Psychology Review*. Vol. 24, pp. 205-249.

- EURYDICE (2009). *National Testing of Pupils in Europe: Objectives, Organization and Use of Results*. Consultado em maio 2010 em <http://www.eurydice.org>.
- EURYDICE (2012). *Key Data on Education in Europe 2012*. Bruxelas: EACEA P9 Eurydice.
- Fernandes, D. (2009). «Avaliação das aprendizagens em Portugal: investigação e teoria da actividade». *Sísifo. Revista de Ciências da Educação*, 09, pp. 87-100. Consultado em abril 2010 em <http://sisifo.fpce.ul.pt>.
- GAVE/JNE. *Resultados das provas de aferição, das provas finais/exames finais nacionais, 2007-2012*.
- GAVE. *Relatórios Nacionais das Provas de Aferição, 2008, 2009, 2010 e 2011*. Consultado em julho 2012 em <http://www.gave.min-edu.pt/np3/24.html>.
- GAVE. *Relatórios Nacionais dos Exames Nacionais, 2009, 2010 e 2011*. Consultado em julho 2012 em <http://www.gave.min-edu.pt/np3/24.html>.
- GAVE. *Relatórios Nacionais dos Testes Intermédios, 2009, 2010 e 2011*. Consultado em julho 2012 em <http://www.gave.min-edu.pt/np3/24.html>.
- Harlen, W. (2006). On the Relationship between Assessment for Formative and Summative Purposes. In J. Gardner (ed.), *Assessment and Learning*. Londres: Sage, pp. 103-118.
- Heritage, M. (2009). «Assessment for Teaching and Learning». Consultado em março 2010 <http://www.k12center.org/publications.html>.
- Hopfenbeck, T. (2011). “Fostering self-regulated learners in a community of quality assessment practices”. *CADMO, Giornale Italiano di Pedagogia Sperimentale*. Ano XIX, 1, 2011, pp.7-20.
- Kellaghan, T., Madaus, G. (2003). External (Public) Examinations. In T. Kellaghan & D. Stufflebeam (eds.), *International Handbook of Educational Evaluation*. Dordrecht: Kluwer Academic Publishers.
- Lambert, D., Lines, D. (2000). *Understanding Assessment, Purposes, Perceptions, Practice*. Nova Iorque: RoutledgeFalmer.
- Lambert, D., Lines, D. (2000). *Understanding Assessment: Purpose, Perceptions, Practice*. Abingdon: RoutledgeFalmer.
- Leong, C. (2006). «On varying the difficulty of test items». *32nd Annual Conference of the International Association for Educational Assessment*. Singapura.
- Looney, J. (2011). «Integrating formative and summative assessment: Progress toward a seamless system?». *OECD Education Working Papers*, n.º 58. Paris: OCDE.

- Martins, C. (2008). *Dez anos de investigação em avaliação das aprendizagens: reflexões a partir da análise de dissertações de mestrado*. Mestrado em Ciências da Educação. Universidade de Lisboa: Faculdade de Psicologia e de Ciências da Educação.
- Mons, N. (2009). «Theoretical and real effects of standard assessment». Bruxelas: EACEA P9 Eurydice.
- Morten R. (2010). «Using student test results for accountability and improvement: a literature review». *OECD Education Working Papers*, n.º 54.
- Nóvoa, A. (2005). *Evidentemente. Histórias da Educação*. Porto: Asa Ed., SA.
- Pellegrino, J. (2006). *Rethinking and Redesigning Curriculum, Instruction an Assessment: What Contemporary Research and Theory Suggests*. Consultado em janeiro 2009 em http://www.skillscommission.org/pdf/commissioned_papers/Rethinking%20and%20Redesigning.pdf.
- Salmerón, L. (2011). ¿Por qué realizar un examen mejora nuestro aprendizaje? Lecciones científicas y educativas del efecto del test. *Ciencia Cognitiva*, 5:1, pp. 19-21. Consultado em maio 2011 em <http://medina-psicologia.ugr.es/cienciacognitiva/?p=228>.
- Santiago, P., Donaldson, G., Looney, A. e Nushe, D. (2012). *OECD Reviews of Evaluation and Assessment in Education: Portugal*. Paris: OCDE.
- Stobart, G (2008). *Testing Time: The uses and abuses of assessment*. Abingdon: RoutledgeFalmer.
- Tindal, G. & Haladyna, T. (eds.) (2002). *Large-Scale Assessment Programs for All Students: Validity, Technical Adequacy, and Implementation*. Londres: Lawrence Erlbaum Associates, Publishers.

Hélder de Sousa é licenciado em Geografia pela Faculdade de Letras da Universidade de Lisboa (1980), professor do ensino secundário desde 1978, leccionou diversas disciplinas na área da formação inicial – Geografia, Ciências do Ambiente, Introdução ao Desenvolvimento Económico e Social (IDES). Director de Serviços de Exames do GAVE, de 2004 a 2010, e director do GAVE desde Março de 2010. Coordenador dos relatórios nacionais de Testes Intermédios e de Exames Nacionais de 2009 a 2011. Co-autor de manuais escolares de Geografia dos ensinos básico e secundário.

QUESTÕES-CHAVE
DA EDUCAÇÃO

AVALIAÇÃO DOS ALUNOS: COMBINANDO AS RAZÕES E OS MODOS

Leandro S. Almeida

AVALIAÇÃO DOS ALUNOS: COMBINANDO AS RAZÕES E OS MODOS¹³

Leandro S. Almeida

Introdução

A avaliação dos conhecimentos, não se confundindo com o ensino e a aprendizagem, informa sobre cada uma destas atividades e a sua interação. Neste sentido, a avaliação dos alunos serve vários objetivos, podendo destacar-se uns ou outros de acordo com os contextos em que se realiza essa avaliação. Pensando-se num contínuo, os objetivos da avaliação podem situar-se por referência a dois polos extremados: (i) a avaliação servindo a certificação das aquisições e a classificação dos alunos nessa base, e (ii) a avaliação servindo a regulação dos processos de ensino e aprendizagem. Pondo-nos em algum ponto deste contínuo, a avaliação dos alunos terá de se diferenciar nos procedimentos e nos instrumentos usados. Não é possível atingir determinados objetivos sem uma atenção às condições e formas da própria avaliação, importando então capacitar os professores e outros técnicos envolvidos

¹³ O autor escrever segundo o novo acordo ortográfico.

numa metodologia apropriada a tal avaliação. Por vezes, atribuem-se à avaliação dos alunos objetivos bastantes gerais como “informar do cumprimento dos programas” ou “ajudar os alunos a aprender”, não se especificando suficientemente a avaliação em causa. Discute-se aqui sumariamente a necessidade de uma explicitação dos objetivos da avaliação dos alunos e as conseqüentes implicações nos seus formatos, ilustrando a necessária dependência entre uns e outros.

Podemos entender a avaliação de conhecimentos dos alunos como um importante momento de recolha de informação sobre os processos de ensino e de aprendizagem, e em particular desta última. A qualidade do ensino do professor, por um lado, e da aprendizagem do aluno, por outro, podem inferir-se através daquilo que o aluno consegue realizar numa determinada situação de avaliação. Neste sentido, a avaliação assume grande relevância para a intervenção escolar e para a decisão política, a qual se pode também comprovar pelo debate frequente no quotidiano escolar. No fundo, a avaliação escolar acaba por ser o meio mais expedito de determinar até que ponto se atingem os objetivos educacionais.

O conhecimento daquilo que o aluno concretiza após um período de aprendizagem serve vários objetivos académicos e sociais. Não pretendendo enumerar todos esses objetivos e a sua diferenciação, queremos sobretudo destacar que uma dada avaliação deve ter bem presente os objetivos que a justificam por forma que assuma o formato e o conteúdo mais adequados a tal finalidade (Barreira, 2001). Por exemplo, reconhecemos que os exames do ensino secundário servem, em Portugal, dois objetivos não totalmente coincidentes: a conclusão deste ciclo de estudos e a obtenção de uma classificação para acesso ao ensino superior. Esta situação, mesmo plausível perante a correlação estatística que sempre se encontra entre as classificações num determinado percurso académico e as aprendizagens e classificações na fase seguinte (quando o currículo

se encontra tendencialmente organizado em cascata ou espiral de abrangência, profundidade e complexidade de conhecimentos), pode fazer-nos questionar da justeza desta prática instituída entre nós. Por exemplo, aponta-se que o ensino secundário está refém do uso que é feito das suas avaliações escolares, ensinando os professores e aprendendo os alunos muito em função daquilo que possa majorar a sua média de acesso ao ensino superior, podendo haver outros objetivos da educação secundária que acabem preteridos. Ensinar/aprender para os testes é bastante redutor do ponto de vista da aprendizagem e, em termos práticos, bastante diferente da situação pedagógica desejável em que os testes podem orientar o processo de ensinar/aprender.

A concluir esta introdução, importa referir que este texto decorre essencialmente de reflexões pessoais, desde logo patente nos tópicos escolhidos para a abordagem do tema. Seguramente que a consulta de estudos e reflexões de outros autores nacionais e internacionais além daqueles apresentados na curta lista de referências enriqueceria a reflexão.

Dos objetivos aos formatos da avaliação

A avaliação dos conhecimentos escolares, não se confundindo com o ensino e com a aprendizagem, informa sobre cada uma destas atividades e sobre a sua interação. A avaliação não se justifica por si mesma. Ela deve estar ao serviço do processo de ensino-aprendizagem, e ao serviço do sistema educativo ou da sociedade. A avaliação é instrumental e dar-lhe outro estatuto poderá significar inverter a lógica das coisas. Como instrumental, carece de algum enquadramento e tecnicidade na sua prossecução, exige uma atenção aos processos e contextos em que ocorre pela interferência destes nos resultados atingidos, e, mais importante ainda, reclama uma clareza quanto à informação a

valorizar e devolver aos agentes implicados. Nem sempre este último ponto, ou seja, o do *feedback* dos resultados da avaliação, aparece em nossa opinião devidamente pensado e operacionalizado em função dos diferentes objetivos, implicações e alvos (sistema educativo/sociedade; professor/escola; alunos/família).

Ao serviço do processo de ensino-aprendizagem, a informação recolhida pode informar-nos como os alunos aprendem aquilo que os seus professores ensinam e aquilo que o seu próprio trabalho escolar lhes proporciona, sendo essa informação relevante para moderar o processo de ensino-aprendizagem. Nesta altura, o professor pode perceber se as suas estratégias de ensino são as mais adequadas à aprendizagem dos seus alunos, e os mesmos alunos podem também verificar o grau em que o seu estudo e trabalho escolar atingem os resultados almejados. A avaliação acaba sempre por ser o espelho da dinâmica da escola apreciando resultados atingidos, em relação aos objetivos fixados, às atividades realizadas e aos recursos envolvidos. Nesta altura, monitorando essa dinâmica, a avaliação deve ser capaz de apontar as mudanças necessárias nas práticas escolares.

A avaliação ao serviço do processo de ensino-aprendizagem é mais contínua que final, é menos formal, assume formas diversas em função dos conteúdos da própria aprendizagem, e adapta-se a grupos de alunos ou à turma. É uma avaliação mais dinâmica, mais próxima das situações de ensino e de aprendizagem, não descarta a subjetividade dos intervenientes e, sobretudo, as formas pessoais ou estilos de ensinar e de aprender. Uma boa situação de avaliação com estes propósitos deve explicitar o aprendido, mas igualmente a forma como foi aprendido e em que medida isso é explicado pela interação entre formas de ensinar do professor e estratégias de aprendizagem dos alunos. Esta informação, mais qualitativa que quantitativa, pode obter-se por via

escrita ou oral, numa situação de teste individual ou de trabalho de grupo, sendo tendencialmente mais bem conseguida, aliás, através da combinação de vários procedimentos e instrumentos de avaliação.

A avaliação de conhecimentos ao serviço do sistema educativo e da sociedade tem o mérito de quantificar o que o aluno aprendeu e também como o professor, a escola e o sistema educativo cumprem a sua função social de preparar as camadas populacionais mais jovens, atingindo os objetivos fixados pelo sistema educativo. De acordo com este propósito, a avaliação tem uma função de classificação e de certificação, apoiando-se a sociedade nos resultados obtidos para apreciar o funcionamento das instituições e agentes de ensino em relação aos objetivos fixados e aos financiamentos públicos havidos. Pode tratar-se, no fundo, de uma certa prestação social de contas. Perante este objetivo, muitas vezes alargando ao território nacional as tarefas de avaliação, a avaliação escolar recorre a provas e a situações bastante formais, a itens tecnicamente bem elaborados (por exemplo, à “teoria da resposta ao item” para estimar parâmetros de precisão e validade ou para conhecer eventuais efeitos de enviesamento dos resultados em função de alguns subgrupos de alunos), procurando assegurar a igualdade dos contextos de avaliação, para garantir a comparabilidade dos resultados atingidos pelos alunos.

Do currículo aos conteúdos da avaliação escolar

Pelo exposto, antecipamos que são múltiplos os objetivos da avaliação escolar, como aliás são múltiplas as funções da própria escola. Tomando o relatório da UNESCO sobre os pilares de uma educação para o presente milénio (UNESCO, 1996), logo aí podemos questionar-nos sobre como avaliar as quatro aprendizagens centrais: aprender a aprender (saber), aprender a fazer, aprender a ser, e aprender a viver

com os outros. Apesar de a generalidade da avaliação escolar se reportar às duas primeiras aprendizagens, a sociedade questiona-se sobre o real impacto da educação formal no desenvolvimento das outras duas aprendizagens. Não pode a escola responder por problemas sociais no seu seio e fora dele quando as competências nestas duas outras áreas não são intencionalmente desenvolvidas.

Na organização do conteúdo e formato das provas (neste caso das situações, questões ou itens das provas) os professores – ou os técnicos que as elaboram – podem orientar-se pelos guiões que acompanham os manuais ou os objetivos de um determinado ciclo de escolaridade. Algumas taxonomias dos objetivos ou funções cognitivas a desenvolver com a aprendizagem podem também orientar a elaboração dos exames, aliás são taxonomias bastante difundidas no mundo da educação. Assim, facilmente se referencia a taxonomia de objetivos de Bloom (1956) a este propósito ou a tipologia de conhecimentos de Alexander e Judy (1988).

Reportando-nos à taxonomia de Bloom, a aprendizagem e a sua avaliação podem assumir ou combinar dois níveis sequenciais de objetivos cognitivos: por um lado, os objetivos reportados à aquisição, diferenciação e organização de conceitos (conhecer, compreender, fixar) e, por outro, objetivos mais centrados no manuseamento e aplicação, seja mais convergente ou mais divergente, da informação (aplicar, analisar, sintetizar, avaliar, criar). Por sua vez, aproximando cognição e conhecimento, Alexander e Judy (1988) tipificam três tipos de conhecimentos que podem organizar a aprendizagem e a sua avaliação: conhecimento factual de tipo declarativo, conhecimento estratégico de índole procedimental ou processual, e conhecimento relativista ou condicional tomando em considerações os contextos e as circunstâncias.

Tabela 1. Objetivos cognitivos do currículo na taxonomia de Bloom

Níveis	Capacidades
conhecer	definir, descrever, distinguir, evocar, recordar ou reconhecer conceitos, factos ou regras
compreender	entender, organizar, classificar, explicar, situar, interpretar ou resumir conceitos ou significados
aplicar	utilizar, demonstrar, interpretar, operar ou resolver situações na base dos conhecimentos
analisar	identificar partes, discriminar, examinar, apreender ou esquematizar relações
sintetizar	combinar partes ou compor um todo coerente, estruturar, desenvolver ou projetar
avaliar	julgar ou apreciar, criticar, defender, escolher e justificar na base de objetivos ou princípios
criar(*)	gerar ou diversificar aplicações, antecipar problemas, produzir nova informação

(*) acrescentado na revisão de Anderson & Krathwohl (2001)

Destacando as situações de avaliação escolar como instrumento pedagógico de diagnóstico e perspetivação das práticas de ensino e de aprendizagem, podem as questões implicar aquilo que o aluno não sabe e não faz, aquilo que o aluno consegue saber e fazer com alguma ajuda ou orientação, e aquilo que sabe e consegue fazer de forma autónoma e fluente. A informação assim recolhida serve não apenas de orientação ao professor, mas também ao aluno, aos encarregados de educação e a outros técnicos de apoio educativo. Por exemplo, os

métodos de estudo, os processos cognitivos e autorregulatórios usados ou, ainda, as abordagens dos alunos à sua aprendizagem, em termos cognitivos e motivacionais, podem não ir além de uma aprendizagem superficial de tipo memorístico, ficando bastante aquém de uma aprendizagem compreensiva ou profunda. A aprendizagem do aluno, também em associação com as práticas de ensino do professor num dado momento, pode centrar-se na aquisição, fluência, manutenção e generalização, podendo esta mesma sequência ser usada na avaliação dos alunos para inferir da qualidade das suas aprendizagens.

Exigências dos instrumentos usados

Podem ser diversos os instrumentos usados na avaliação escolar. Quando esta avaliação classifica, é dominante internacionalmente o recurso a provas ou a exames. Por comodidade, tomaremos estes dois termos como sinónimos. As provas para uma avaliação destinada à classificação e certificação dos alunos, tomados no seu todo nacional, carecem de três características metrológicas importantes: sensibilidade, fidedignidade e validade (Almeida & Freire, 2010). Podemos ilustrar estas características tomando, por exemplo, os conhecidos exames nacionais. A sensibilidade prende-se com a capacidade de os resultados obtidos traduzirem os níveis diferenciados de capacitação (conhecimentos, informação, competências) existentes nos próprios alunos. Se o objetivo é classificar, mais ainda quando os exames servem o objetivo de seriar os alunos no seu acesso ao ensino superior, então as provas devem ser capazes de diferenciar tais alunos em função da profundidade, quantidade ou complexidade dos seus conhecimentos. Não se trata de ser a prova a produzir as diferenças (isso seria um teste mal construído), mas de uma prova ser capaz de captar as diferenças relevantes nos conhecimentos

dos alunos face aos programas ou aos objetivos de uma determinada avaliação. A sensibilidade de uma prova traduz que dois alunos com o mesmo nível de conhecimento não se diferenciam na classificação atingida; contudo será tanto mais sensível quanto for capaz de diferenciar, nos níveis de desempenho, competências ligeiramente afastadas. Por vezes, recorre-se a uma distribuição gaussiana (curva normal de distribuição dos resultados) para demonstrar a sensibilidade de um instrumento de avaliação. Nessa distribuição, uma larga maioria dos sujeitos pontua em torno da média, diminuindo progressivamente a sua frequência à medida que nos afastamos da média (por norma, há valores até três unidades de desvio-padrão acima e abaixo da média). Importa referir que este conceito, podendo ser importante em provas normativas (testes referenciados à norma ou *norm-referenced tests*), não faz sentido nas situações de avaliação escolar quando o objetivo é informar das aprendizagens dos alunos para um percurso de ensino-aprendizagem que se deseja estruturar. Nessa altura, numa fase inicial da aprendizagem poderíamos antecipar que a totalidade dos alunos apresenta fracas conhecimentos e competências na área (embora importando sempre saber quais são os conhecimentos e as competências possuídas, e em particular os pré-requisitos), sendo desejável que no final do período da aprendizagem a avaliação mostre uma inversão na curva de distribuição dos valores, havendo poucos alunos sem as aprendizagens adquiridas e a quase totalidade dos alunos com as mesmas efetuadas. Falamos, neste caso, de uma avaliação mais de critério e centrada na verificação da mestria (testes referenciados a critério ou *criterion-referenced tests*). Um aprofundamento da diferenciação entre provas referenciadas a normas e a critério pode encontrar-se em Ribeiro e Almeida (2001), onde se desenvolvem os procedimentos metodológicos a considerar na construção e validação de provas referenciadas a critério.

A fidedignidade ou precisão de uma prova reflete-se na confiança que podemos ter nos seus resultados. Um termómetro preciso dá-nos confiança sobre a temperatura do corpo e um eventual estado febril. Testamos isso, por exemplo, quando voltamos a usar o termómetro (método teste-reteste) e verificamos que a temperatura do corpo é a mesma. A margem de erro sempre presente em qualquer mensuração é relativamente reduzida quando a prova é precisa, entendendo-se então que os resultados traduzem as diferenças individuais realmente existentes no construto ou domínio que se está a avaliar. Quando uma prova de um aluno obtém uma classificação diferente ao passar pelas mãos de um segundo avaliador, podemos inferir que existem problemas de precisão da avaliação (da prova ou da sua grelha de correção, partindo-se do pressuposto de que ambos os avaliadores estão devidamente creditados para essa avaliação). Mesmo aceitando a presença óbvia de fatores de erro em toda a avaliação escolar envolvendo o julgamento do professor, designadamente em questões abertas de resposta direta ou desenvolvida que requerem a apreciação do professor sobre a sua adequação (Pacheco, 1998), é importante reduzir o impacto de tal subjetividade nos resultados. No dia em que pais e alunos reclamem da ponderação do erro-padrão da medida na reapreciação das suas classificações, provavelmente o país terá mais um problema em mãos, relativo à seriação dos alunos para responder à política de *numerus clausus* no acesso ao ensino superior.

Finalmente, a validade da prova remete-nos para o valor da informação assim recolhida. Desde logo, interessa ver se a prova cobre adequadamente os conteúdos e objetivos do programa e se valoriza, através das questões formuladas e da sua forma de pontuação, o mais importante do currículo (validade facial e validade de conteúdo). A validade da prova inclui, ainda, o sentido prático da informação. Por exemplo,

acreditamos que um aluno com uma classificação de 18 num exame nacional de Biologia e Geologia tem melhor preparação para aprender Biologia ou Geologia num nível seguinte, em comparação com um colega que apenas obteve uma classificação de 15 nesse exame. Se tal não ocorre, então o exame avalia outras coisas que não os conhecimentos ou competências dos alunos num determinado currículo de Biologia e Geologia. Se as classificações com que os alunos acedem ao ensino superior pouco dizem sobre os níveis de sucesso atingidos nas mesmas áreas científicas no final do 1.º ano, então a pergunta legítima é a de saber por que razão usar tais classificações para seriar os alunos no acesso ao ensino superior público em Portugal. Da mesma forma, e saindo da escola para a sociedade, se os gestores de recursos humanos acreditam piamente que as competências profissionais de um diplomado pouco ou nada têm que ver com a média final da sua graduação, poder-se-á questionar a razão dessa pergunta numa entrevista de emprego. De novo, os objetivos prosseguidos com a avaliação definem o tipo de prova e, conseqüentemente, o tipo de informação recolhida. Precisamos, como país, de clarificar o que pretendemos com as avaliações em curso para melhor delimitarmos o conteúdo e formato dessas avaliações. A título de exemplo, entrando numa *avaliação formativa* por contraponto a uma *avaliação sumativa*, antecipamos a pertinência de uma avaliação recorrendo à construção e análise de portefólios individuais ou de pequenos grupos de alunos, aliás um instrumento de avaliação que é consonante com uma conceção construtivista de aprendizagem, assente no papel ativo do aluno na construção de conhecimento e no desenvolvimento de competências.

Por vezes, atribuem-se à avaliação dos alunos objetivos bastantes genéricos, tendencialmente benévolos e supostamente inofensivos, como “informar do cumprimento dos programas” ou “ajudar os

alunos a aprender”, não se especificando suficientemente a avaliação em causa. Em nossa opinião, é fundamental explicitar, sem rodeios, os objetivos de uma dada avaliação escolar dos alunos dadas as implicações ou interdependência com os formatos dessa mesma avaliação. Sendo legítimo à sociedade e aos seus governos efectuarem avaliações para controlo do funcionamento do sistema educativo no seu todo e das escolas, pode ser já questionada a forma como tal avaliação é feita e como os resultados são publicitados, por vezes dando origem a *rankings* de escolas de que uma opinião pública, sem qualquer formação relativa aos contextos e processos inerentes, faz uso. Os exames, certificando o volume e a qualidade das aprendizagens, não fazem aumentar *per si* a qualidade da aprendizagem. A aprendizagem depende do trabalho dos professores, dos alunos, dos encarregados de educação, dos governos e da sociedade em geral. Punhamo-nos de acordo e vejamos como cada um pode contribuir para este objetivo nacional!

Considerações finais

Para terminar, e como síntese, destacamos algumas ideias. Em primeiro lugar, a avaliação escolar prossegue diversos objetivos. Importa, por isso, ter estes objetivos claros, conceber a avaliação em função dos mesmos e ser parcimonioso na prossecução das atividades de avaliação. A avaliação, em si mesma, não tem qualquer efeito miraculoso na melhoria do sistema educativo, das práticas escolares e das aprendizagens dos alunos. De qualquer modo a avaliação e o recurso a instrumentos apropriados para a sua prossecução estão presentes nos sistemas educativos apontados como mais “produtivos”, assim como nenhuma reforma educativa prescinde da referência explícita à avaliação escolar.

Em segundo lugar, qualquer avaliação escolar deve recorrer a procedimentos e instrumentos adequados aos seus objetivos. Os conteúdos versados e os formatos dos instrumentos não são alheios aos objetivos de uma avaliação concreta. Mesmo as questões fechadas podem diferenciar-se na informação recolhida consoante os elementos informativos necessários à resposta estejam presentes, cabendo ao aluno o seu *reconhecimento*, ou ausentes, exigindo do aluno já a sua *evocação*.

Em terceiro lugar, os objetivos prosseguidos por uma dada avaliação devem orientar a leitura dos resultados obtidos. Provavelmente esta preocupação ajuda-nos a precisar o que, quando e como avaliar. Sendo a avaliação instrumental, torna-se importante cuidar da forma como o *feedback* dessa avaliação é comunicada e trabalhada pelos vários intervenientes e envolvidos, sendo certo que a par dos professores e alunos, a avaliação dos alunos serve o sistema educativo e a sociedade, a escola e a família. Reconheceremos que todos eles processam essa informação de forma diferente, diferindo desde logo no tipo de informação de que necessitam.

Acrescentaríamos agora dois últimos pontos na avaliação escolar. Em primeiro lugar, a importância do avaliador e/ou do utilizador da informação recolhida. A expectativa é que os profissionais mais diretamente implicados sejam melhores (competentes) que os testes que usam, por outras palavras não podemos pôr o ónus no instrumento, pois este carece sempre de mãos sábias, seja na sua construção seja na sua rentabilização. Precisamos de evoluir na qualidade dos instrumentos de avaliação. Aliás, esse esforço tem ocorrido em Portugal nos exames nacionais, cuja construção e validação recorre às metodologias mais avançadas de análise estatística, em particular a procedimentos de análise enquadrados na “teoria da resposta ao item”, baseando a informação recolhida mais nos construtos avaliados do que nas

subamostras de sujeitos que respondem aos testes (Primi & Almeida, 2001). Não podemos, porém, descurar a formação dos técnicos envolvidos nas fases sucessivas do processo avaliativo.

Em segundo lugar, uma referência aos alunos, os sujeitos da avaliação escolar quando esta avaliação se dirige à verificação das suas aquisições. A avaliação é sempre uma situação de alguma complexidade, vivenciada em geral com alguma ansiedade. O desempenho acadêmico dos alunos nas situações de avaliação reflete reconhecidamente os seus conhecimentos e competências, mas também o contexto e o processo da própria avaliação, e mais ainda outras características pessoais que vão além do objeto da avaliação. Neste caso, vários estudos na área da psicologia realçam os estilos cognitivos dos alunos que podem interferir nas suas respostas em função dos formatos das questões (por exemplo, o estilo da dependência *versus* independência de campo ou da produção divergente *versus* produção convergente), as suas percepções de eficácia, as capacidades de resiliência, de controlo e de gestão de *stress*, os níveis de motivação académica ou, ainda, os níveis de ansiedade (podendo, consoante a intensidade, serem estimulantes ou debilitantes do desempenho), entre outras. Neste campo temos realizado alguma investigação sobre o impacto das *atribuições causais* no rendimento académico, associando-as às explicações que os alunos apresentam para os seus bons ou fracos resultados nas avaliações. Alguns alunos, por exemplo, apresentam padrões atribucionais pouco adequados justificando os seus maus resultados nos testes com a falta de capacidades cognitivas (causa percebida como interna, estável e não modificável), o que, com alguma frequência, os leva a situações de desânimo e de abandono das tarefas escolares nas disciplinas em causa. Os estudos apontam a necessidade de os professores estarem atentos a esta situação, ajudando tais alunos a reelaborarem as suas cognições e a

desfocarem tais atribuições das suas aptidões intelectuais (Barros & Almeida, 1991; Almeida, Miranda, & Guisande, 2008).

Não se deve descurar o impacto deste conjunto de fatores na avaliação, designadamente nos níveis, quantitativos e qualitativos, de prestação. Aliás, alguns destes construtos pessoais do aluno são psicologicamente constituídos em decorrência das avaliações ao longo do seu percurso escolar e da forma como ele foi processando o respectivo retorno.

Referências bibliográficas

- Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research*, 58, 375-405.
- Almeida, L. S., & Freire, T. (2010). *Metodologia de investigação em psicologia e educação* (5.ª edição). Braga: Psiquilíbrios.
- Almeida, L. S., Miranda, L., & Guisande, M. A. (2008). Atribuições causais para o sucesso e fracasso escolares. *Estudos de Psicologia*, 25(2), 169-176.
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing*. New York: Longman.
- Barreira, C. M. F. (2001). Duas estratégias complementares para a avaliação das aprendizagens: A avaliação formadora e a avaliação autêntica. *Revista Portuguesa de Pedagogia*, 35 (3), 3-33.
- Barros, A. M., & Almeida, L. S. (1991). Dimensões socio-cognitivas do desempenho escolar. In L. S. Almeida (Ed.), *Cognição e aprendizagem escolar*. Porto: Associação dos Psicólogos Portugueses (APPORT).
- Bloom, B. S. (Ed.) (1956). *Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain*. New York: Longman.
- Pacheco, J. A. (1998). Avaliação da aprendizagem (pp. 111-132). In L. Almeida & J. Tavares (Orgs.), *Conhecer, aprender e avaliar*. Porto: Porto Editora.
- Primi, R., & Almeida, L. S. (2001). Teoria da resposta ao item (TRI). In E. M. Fernandes & L. S. Almeida (Eds.), *Métodos e técnicas de avaliação: Contributos para a prática e investigação psicológicas*. Braga: Universidade do Minho, Centro de Estudos em Educação e Psicologia.

Ribeiro, I. S., & Almeida, L. S. (2001). Testes referenciados a critério. In E. M. Fernandes & L. S. Almeida (Eds.), *Métodos e técnicas de avaliação: Contributos para a prática e investigação psicológicas*. Braga: Universidade do Minho, Centro de Estudos em Educação e Psicologia.

UNESCO (1996). *Educação, um tesouro a descobrir. Relatório para a Unesco da Comissão Internacional sobre a Educação para o século XXI*. Porto: Asa. (Relatório coordenado por Jacques Delors).

Leandro S. Almeida é Professor catedrático do Departamento de Psicologia da Educação e Educação Especial do Instituto de Educação da Universidade do Minho. Doutorou-se em Psicologia, especialidade de Psicologia da Educação, na Universidade do Porto, em 1987, e prestou provas de agregação em Psicologia na Universidade do Minho em 1994. É autor de duas centenas de trabalhos, entre livros, capítulos de livros e artigos, publicados em Portugal e no estrangeiro.

Foi director do Centro de Estudos em Educação e Psicologia da Universidade do Minho (1997/99); vice-presidente do Conselho Académico (2000/02) e vice-reitor (2006/09); sendo actualmente presidente do Instituto de Educação.