

F O L H A #

Para uma abordagem pedagógica dos testes

Eusébio André Machado

Março de 2021



PROJETO DE
MONITORIZAÇÃO
ACOMPANHAMENTO
E INVESTIGAÇÃO
EM AVALIAÇÃO
PEDAGÓGICA



FOLHA

Para uma abordagem pedagógica dos testes

Eusébio André Machado

Março de 2021



Cofinanciado por:



Fundo Social Europeu



Índice

Introdução	3
Tipologia do uso dos testes: classificar e avaliar	6
Para uma abordagem pedagógica dos testes	9
Tarefas	15
Referências	16

Introdução

No âmbito da educação, desde meados da primeira década do século XX, os testes têm sido objeto de uma preocupação permanente e quase dominante de natureza psicométrica: *como medir objetivamente as aprendizagens?* Os pioneiros da introdução “testes” na educação, sob a influência do positivismo reinante, pretendiam contruir uma ciência objetiva para medir os resultados das aprendizagens dos alunos (Giordano, 2007, p. 29), à luz, de resto, do que então ocorria no domínio da psicologia.

Na origem da necessidade de uma ciência dos testes (mais tarde, conhecida como “docimologia”), para além de razões de natureza epistemológica, está a prevalência de um paradigma de avaliação educacional, segundo o qual *avaliar é classificar*. Neste sentido, devido a uma preocupação de medição objetiva para comparar, selecionar e hierarquizar os alunos (veja-se a Figura 1), o *uso pedagógico* dos testes tem sido frequentemente negligenciado, sobretudo no âmbito de uma avaliação *para* as aprendizagens.



Figura 1. O novo método de medir através dos testes (Reimpresso do *American School Board Journal*, abril 1922)



Ora, não se pretende com esta *Folha* menosprezar a importância das questões psicométricas relativas à construção dos testes, até porque, quanto melhor for a qualidade da informação recolhida, melhor se podem tornar os processos de apoio às aprendizagens: quer se trate da avaliação sumativa, designadamente com fins classificatórios, quer se trate da avaliação formativa, a qualidade psicométrica dos processos de recolha de informação é uma condição instrumental para o sucesso do ensino e da aprendizagem.

Convém enfatizar, porém, que os testes usados em contexto escolar e no âmbito da avaliação interna possuem uma “validade pedagógica” porque são concebidos em função de um elevado grau de contextualização, adaptação e individualização, aplicando-se a um conjunto relativamente reduzido de alunos (uma turma, um grupo de alunos ou até a um aluno), ao contrário dos chamados “testes estandardizados” elaborados no âmbito da avaliação externa. Neste sentido, os professores, nas escolas, por razões óbvias, não têm possibilidades de garantir a validade, a fiabilidade, o nível de discriminação e o índice de dificuldade dos testes, as quatro “medidas” psicométricas por excelência.

Assim, sob o paradigma de uma avaliação orientada para melhorar os processos de ensino e aprendizagem – *e não apenas para classificar* -, entendemos que importará enfatizar *uma abordagem pedagógica dos testes*. Trata-se de tentar perceber de que modo os testes se podem constituir como processos de recolha de informação ao serviço das aprendizagens, da avaliação formativa e da distribuição de um *feedback* de elevada qualidade e efetividade. Dito de outro modo, trata-se de colocar os testes ao serviço de uma pedagogia que Cardinet (1993, p. 23) resume do seguinte modo: *facilitar a aprendizagem* dos alunos.

Convirá lembrar que os testes são dos processos de recolha de informação em educação mais bem-sucedidos e eficazes, não obstante as limitações e problemas identificados pela literatura (e.g., Black & Wiliam, 1998), permitindo aos professores e aos alunos obter, em



qualidade e em quantidade, dados extremamente relevantes para regular o ensino e a aprendizagem.

Ao longo dos últimos cem anos, a investigação tem dado relevo ao chamado “efeito-teste” (*testing effect*), sendo que, como referem Eisenkraemer, Jaeger e Stein (2013), apoiados numa revisão sistemática de literatura, há evidências empíricas segundo as quais a aplicação repetida de testes, *até um certo ponto*, durante a aprendizagem, não só fornece o conhecimento de resultados com implicação direta na aprendizagem, como também favorece fortemente a retenção da informação a longo prazo. Na realidade, apesar de ser uma palavra negativa para muitos professores e alunos, “os testes podem ser uma das mais potentes ferramentas para melhorar a aprendizagem” (Rawson & Dunlosky, 2012, p. 420). A questão que se coloca é, pois, saber de que forma o uso de testes pode beneficiar as aprendizagens dos alunos no âmbito de uma avaliação pedagógica (Fernandes, 2019).



Tipologia do uso dos testes: classificar e avaliar

Os testes constituem um processo de recolha de informação com elevado grau de formalização e de complexidade no que diz respeito, principalmente, às suas *lógicas de construção, modos de administração e processos de restituição*, pelo que exigem um conjunto de opções que devem estar de acordo com as *finalidades de avaliação* pretendidas. Em função destas variáveis de caracterização, apresentamos a seguir uma tipologia (Figura 2) que identifica os dois principais tipos de testes em função dos seus usos em contexto escolar: *os testes usados para classificar e os testes usados para avaliar*¹.

Variáveis de caracterização	Testes usados para <i>classificar</i>	Testes usados para <i>avaliar</i>
Lógicas de construção	Tendo como objetivo a classificação ordenada dos indivíduos, são construídos para obter a maior diferenciação possível entre indivíduos em função de uma distribuição normativa e incidem sobre um número significativo de objetivos de aprendizagem	Tendo como propósito apoiar os alunos, são construídos para compreender o melhor possível as dificuldades de aprendizagem em função de critérios definidos e incide sobre um número limitado ou reduzido de objetivos de aprendizagem
Modos de administração	São aplicados invariavelmente a um grupo de alunos (turma/classe) de forma simultânea e sob as mesmas condições para garantir um princípio de igualdade (modelo do exame)	São aplicados em grupo ou individualmente, de forma simultânea ou não, permitindo diferentes condições de realização para garantir um princípio de equidade
Processos de restituição	A restituição da informação realiza-se sob a forma de uma classificação, de natureza fortemente normativa, centrada no ensino e que integra o desempenho do aluno em relação a todos os objetivos de aprendizagem	A restituição da informação realiza-se sob a forma de <i>feedback</i> (oral ou escrito, com uma natureza acentuadamente criterial, centrada nas aprendizagens e discriminando o desempenho em relação a cada um dos objetivos de aprendizagem

¹ Uma tipologia mais frequente, quer na literatura, quer nas práticas de avaliação, assenta na distinção entre *testes formativos e testes sumativos*. Neste caso, adotamos os princípios da tipologia proposta por Cardinet (1993), na qual surgem vários tipos de testes: *testes de classificação, testes de interesses, testes de dificuldades crescentes, testes de fatores*, etc. A nossa principal justificação para esta tipologia é a necessidade de separar os processos de recolha de informação das finalidades da avaliação, o que, no caso dos testes, tem constituído um obstáculo difícil de ultrapassar.



Finalidades de avaliação	Avaliação sumativa com fins classificatórios, permitindo, no limite, selecionar, comparar e/ou hierarquizar, embora também possam ter um uso formativo	Avaliação formativa assente na distribuição de <i>feedback</i> de alta qualidade, individualizado e em função de critérios partilhados com os alunos
---------------------------------	--	--

Figura 2. Uma tipologia do uso dos testes

Embora se trate, como em todas as tipologias, de uma tentativa de explicitação abstrata de uma realidade altamente complexa, entendemos que, não obstante, pelas suas eventuais capacidades heurísticas, esta tipologia pode contribuir para dois objetivos fundamentais no que à avaliação diz respeito: por um lado, mostrar que é possível, como veremos adiante, combinar de diferentes formas as variáveis de caracterização dos testes, sobretudo no que se refere aos testes usados para avaliar, pondo em causa a forte tradição do modelo único de teste; por outro lado, reiterar o princípio geral defendido por Cardinet, segundo o qual os testes devem estar adaptados às finalidades de avaliação pretendidas para se maximizar as suas potencialidades pedagógicas. Neste aspeto, vale a pena citar as palavras do autor:

Não existem bons testes em si mesmos, nem uma avaliação pedagógica modelo, como os tratados de docimologia têm tendência a fazer crer. Os instrumentos de medida terão de ter características diferentes e, muitas vezes, opostas, conforme a finalidade a atingir e segundo o tipo de decisão a tomar. Querer obter tudo duma vez é condenar-se a não saber aquilo que se obtém. (Cardinet, 1993, p. 70)

As práticas de avaliação dominantes em contexto escolar assentam no uso de um único tipo de testes (os testes usados para classificar) que, nas suas opções fundamentais, segue o modelo do exame e da avaliação externa: é concebido com características psicométricas que procuram discriminar ao máximo as diferenças entre os alunos, são administrados de modo coletivo, simultâneo e uniforme e são restituídos sob a forma de uma classificação (e/ou através de uma menção qualitativa) de acordo com a escala adotada. De resto, este tipo de testes, com a ilusão de uma alegada objetividade, acaba por ser reforçado por sistemas educativos que sobrevalorizam os exames nacionais e que incutem um carácter altamente seletivo à escola e, por inerência, às práticas de avaliação, dando origem ao fenómeno de “ensinar para os testes”.



No entanto, é fundamental reforçar a ideia de que o que está em causa não são os testes em si, mas o facto de se verificar que praticamente estão condenados a um único uso: classificar. Por outro lado, importa também salientar o uso pedagógico dos testes, independentemente da finalidade (classificar ou avaliar) e numa lógica de complementaridade entre a avaliação sumativa e a avaliação formativa.



Para uma abordagem pedagógica dos testes

No que diz respeito aos *testes concebidos e aplicados com uma finalidade classificatória*, no âmbito de uma avaliação sumativa, podem ter um uso pedagógico que reforce e melhore a qualidade das aprendizagens dos alunos, ou seja, podem ter um uso formativo. Convém lembrar, contudo, que este tipo de testes tem por objeto um conjunto significativo de aprendizagens (podem ser considerados *testes de banda larga*) e é aplicado, por consequência, após a conclusão de uma unidade, de um módulo ou de um período letivo. Dado este carácter *a posteriori*, que implica uma separação entre o currículo e a avaliação, surgem várias limitações para o uso formativo e pedagógico deste tipo de testes, designadamente a regulação das aprendizagens, o *feedback* oportuno e específico e a oportunidade para a melhoria. Não obstante, há um conjunto de estratégias de natureza pedagógica que é possível associar ao uso de testes com finalidade classificatória e que se enquadram naquilo que se designa por um “uso formativo dos testes sumativos” (Hopster-den Ottera, Woosb, Eggena, & Veldkampc, 2017; Lam, 2012).

A este propósito, Black, Harrison, Lee, Marshall e Wiliam (2011, pp. 53-56) propõem três “inovações”, tendo como propósito mais abrangente, em termos de políticas de avaliação, mostrar que os chamados “testes sumativos” devem ser, e devem ser vistos, como um elemento positivo do processo de aprendizagem.

A primeira inovação consiste em usar práticas formativas no âmbito das revisões realizadas antes dos testes. Neste caso, trata-se de tornar as estratégias de revisão mais efetivas, através das quais os alunos assumem um papel ativo, planeando a sua própria revisão. É importante que os alunos sejam capazes de identificar os objetivos de aprendizagem nos quais devem concentrar o seu estudo. De seguida, os alunos podem ser convidados a identificar questões em exames e testes que se encontrem no âmbito dos objetivos de aprendizagem em relação aos quais se sentem menos seguros. Por fim, podem recorrer a manuais ou aos pares, designadamente em trabalho de grupo, para garantir que conseguem responder adequadamente às questões selecionadas.

A segunda inovação, ainda na fase preparatória de um teste ou de um exame, reside em solicitar aos alunos que formulem e, depois, respondam às suas próprias questões. Quando



os alunos são convidados a colocar as questões e a classificar as respostas, compreendem melhor o processo de avaliação e orientam o seu estudo de modo mais eficaz. Este compromisso mais ativo dos alunos com as suas aprendizagens, encarados como intervenientes diretos, tem um efeito de responsabilização e de desenvolvimento de competências metacognitivas cuja importância é fundamental em qualquer processo de aprendizagem.

A terceira inovação, incluída na fase pós-teste, corresponde à criação de oportunidades para trabalho de natureza formativa. Desde logo, o professor pode selecionar as questões nas quais os alunos revelaram mais dificuldades e, em vez de concentrar o diálogo nos pontos atribuídos, incidir na revisão das aprendizagens que estão na base das referidas questões com desempenhos mais fracos. Por outro lado, os alunos podem ser convidados a classificar os testes dos seus pares, mormente em trabalho de grupo, recorrendo aos critérios ou à rubrica fornecida pelo professor. Depois, através de um processo de tutoria, os alunos podem ajudar os pares na resolução das questões, deixando para o professor as questões com particular dificuldade.

Uso formativo dos “testes sumativos”: um exemplo

Apresentamos, de seguida, um exemplo de recurso a estratégias normalmente associadas à avaliação formativa e que, neste caso, são aplicadas à realização de um “teste sumativo”. Neste exemplo procuramos combinar as sugestões da “avaliação por contrato de confiança” (Merle, 2018) e das propostas de “uso formativo dos testes sumativos” de Black, Harrison, Lee, Marshall e Wiliam (2011) acima referidas.

1. Antes da realização do teste (uma ou duas semanas) o professor comunica aos alunos os *objetivos de aprendizagem, bem como os critérios de avaliação*, o que configura uma *lógica feed up*. Neste caso, este procedimento pode traduzir-se sob a forma de uma rubrica, cujo valor pedagógico estará diretamente relacionado com a compreensão que os alunos venham a ter da mesma. A informação prévia prestada pelo professor aos alunos assenta num “princípio de alinhamento curricular”, ou seja, num isomorfismo rigoroso e transparente entre os objetivos de aprendizagem e os processos de recolha de informação: a avaliação deve incidir sobre as aprendizagens previstas e efetivas.



2. Ao mesmo tempo e antes do teste, o professor fornece um *programa detalhado de revisão* constituído por um conjunto de atividades, problemas, questões e exercícios, o qual se encontra em relação direta com as aprendizagens realizadas em sala de aula, o que deve permitir ao aluno preparar-se adequadamente para o teste. O professor deve informar e garantir aos alunos que a maior parte dos exercícios do teste incidirão sobre o programa detalhado de revisão proposto.

3. Ainda antes do teste, o professor deve promover momentos de natureza eminentemente pedagógica para a *distribuição de feedback* oportuno, específico e de elevada qualidade em torno do programa de revisão. Deste ponto de vista, este momento deve permitir que os alunos se confrontem com as suas dificuldades e dialoguem com o professor e com os pares; não se trata, pois, de o professor mostrar como resolve o material que forneceu no seu programa de revisão. Os alunos são, pois, incentivados a colocar questões acerca das aprendizagens e dos exercícios sobre os quais têm dificuldades.

4. Após a realização dos testes (sem deixar que os alunos esqueçam o que fizeram), não obstante as limitações que referimos, haverá outro momento de *distribuição de feedback* com base nos critérios de avaliação ou da rubrica fornecidos pelo professor antes do teste. Este *feedback* pode revestir-se de diversas modalidades, podendo ser oral ou escrito, individual ou coletivo, conforme o que for mais adequado e eficaz.

5. Ainda após a realização dos testes, é importante o *feed forward* em função dos critérios definidos, quer para os alunos, quer para os professores. O *feed forward* deve ser um trabalho conjunto, envolvendo professor e alunos, designadamente através de uma revisão pelos pares e de tutorias. É fundamental que os alunos recebam orientações em relação às aprendizagens que necessitam de rever, aos procedimentos que devem adotar e, nos casos de maior insucesso, os apoios que devem receber, sem nunca desresponsabilizar os alunos pelas suas aprendizagens. No caso dos professores, surge a necessidade de adaptação do ensino para que possam ajudar os alunos a superar as dificuldades evidenciadas, diferenciando as estratégias e agindo com elevado grau de especificidade.



Por sua vez, no que respeita ao *uso de testes com a finalidade de avaliar*, sendo certo que resultam de uma intencionalidade formativa, importa considerar, pelo menos, três aspetos que podem contribuir para a sua maior efetividade pedagógica: em primeiro lugar, devem incidir sobre um conjunto reduzido de objetivos de aprendizagem (são *testes de banda estreita*) e, de facto, o seu valor pedagógico depende da sua inserção, *a par e passo*, nos processos de ensino e aprendizagem; em segundo lugar, uma vez que não visam diferenciar, podem ter apenas um tipo de perguntas, visar um único nível taxonómico ou possuir um grau de dificuldade uniforme, sempre em função do propósito de recolher a melhor informação para apoiar os alunos; em terceiro lugar, e em função da finalidade formativa, os testes podem ser administrados e realizados de forma mais flexível: não têm que ser aplicados a todos os alunos, o tempo de realização pode e deve ser diferenciado e mesmo as perguntas podem ser diferentes de aluno para aluno. Em suma, este tipo de testes não segue uma lógica de padronização, como acontece nos testes e nos exames concebidos com a finalidade de classificar.

Mas, se o propósito do teste é a avaliação enquanto processo eminentemente pedagógico, o que está em causa é a qualidade e a pertinência da informação recolhida para a promoção de mais e melhores aprendizagens: o teste deve estar ao serviço das aprendizagens e não são as aprendizagens que devem estar ao serviço do teste. Deste ponto de vista, a *pregnância pedagógica* de um teste usado para avaliar depende principalmente do processo e da qualidade da restituição da informação aos alunos, bem como das oportunidades que são proporcionados para a superação de dificuldades e para a melhoria. Assim, um teste com um propósito formativo é utilizado para efeitos de *feedback*, conforme os processos e características referidos na literatura (Brookhart, 2008) e com um elevado nível de participação e envolvimento dos alunos (Machado, 2020).

Estratégia de *Feedback* para um “teste formativo”: um exemplo

Trata-se de um teste que foi aplicado a um grupo de sete alunos em relação aos quais, durante as aulas, foram identificadas dificuldades na aquisição de um determinado objetivo de aprendizagem. O teste é constituído por dois exercícios relativos ao objetivo



de aprendizagem em causa, sendo que os alunos dispuseram de tempos diferentes para responder (grosso modo, entre 10 a 20 minutos).

1. Antes de entregar o teste, o professor explicita os objetivos de aprendizagem e os critérios de avaliação. Para o efeito, entrega uma rubrica com quatro critérios, indicando três níveis de proficiência em relação ao objetivo de aprendizagem acerca do qual os alunos manifestaram dificuldades.

2. Após a realização do teste, os alunos são convidados a partilhar com os pares, em regime de tutoria, as respostas dadas. Tendo por base a rubrica fornecida, em trabalho com os pares, verificam as principais dificuldades e erros cometidos e discutem as melhores estratégias para a resolução dos dois exercícios.

3. Durante este processo, o professor supervisiona o trabalho que está a ser realizado pelos alunos, intervindo nas questões particularmente difíceis e dando orientações precisas aos alunos. Ao mesmo tempo, recolhe informação para adaptar as suas estratégias de ensino em relação ao referido objetivo de aprendizagem.

4. De seguida, o professor convida novamente os alunos a responder ao teste, eventualmente diminuindo o tempo, embora considerando a possibilidade de haver ritmos diferentes de resposta. É, pois, dada uma nova oportunidade para os alunos verificarem a sua evolução.

5. No final, o professor dialoga com os alunos sobre as dificuldades que tiveram e as estratégias que adotaram. Neste contexto, aproveita para dar um *feedback* oral e dirigido a todos os alunos, sintetizando as principais estratégias a seguir para uma aquisição plena do objetivo de aprendizagem que motivou o teste formativo.

Nos testes usados para avaliar, é muito importante que o *feedback* seja tão oportuno quanto possível, pelo que devem ser implementadas estratégias formativas que permitam aos alunos receber o apoio e as orientações necessárias após a realização do teste, não sendo necessário que, em todas as circunstâncias, o professor recolha e analise os testes. De resto, a adoção da lógica de padronização dos testes que têm por objetivo classificar



acaba por ser um forte dissuasor do recurso mais frequente e sistemático aos testes com propósitos formativos. Na realidade, se os chamados testes formativos são, por norma, aplicados a todos alunos, obrigando o professor a recolher e a analisar todos os testes, não só se perde a oportunidade do *feedback*, como também o tempo despendido dificilmente se torna compatível com a vida profissional dos professores.



Tarefas

As tarefas que se seguem dizem respeito ao *uso pedagógico dos testes* e devem ser realizadas em pequenos grupos.

Tarefa 1.

Discuta e interprete a Figura 1, tendo em conta as práticas de uso dos testes em contexto escolar.

Tarefa 2.

Discuta e justifique a importância pedagógica dos testes.

Tarefa 3.

Elabore, discutindo com os seus colegas uma estratégia formativa para aplicar a um teste concebido para classificar, tendo em conta o exemplo apresentado nesta *Folha*.

Tarefa 4.

Discuta a compatibilidade (ou não) entre *feedback* (realização de comentários orais ou escritos, de natureza eminentemente qualitativa e descritiva, com o intuito de ajudar os alunos) e classificação (atribuição de notas) no âmbito dos testes com finalidades formativas.



Referências

- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5, (1), 7-74.
- Black, P., & Wiliam, D. (2006). Assessment for learning in the classroom. In J. Gardner (Edi.) *Assessment for learning* (pp. 9-25). Sage Publications
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Open University Press.
- Brookhart, S. M. (2008). *How to give effective feedback to your students*. ASCD.
- Cardinet, J. (1993). *Avaliar é medir?* Edições Asa.
- Eisenkraemer, R. E., Jaeger, A., & Stein, L. M. (2013). A systematic review of the testing effect in learning. *Paidéia*, 23, (56), 397-406.
- Fernandes, D. (2019). Para um enquadramento teórico da avaliação formativa e da avaliação sumativa das aprendizagens escolares. In M. I. R. Ortigão, D. Fernandes, T. V. Pereira & L. Santos (Org.), *Avaliar para aprender no Brasil e em Portugal* (pp. 139-163). Editora CRV.
- Giordano, G. (2007). *How testing came to dominate American schools. The history of educational assessment*. Peter Lang.
- Hopster-den Ottera, D., Woolsb, S., Eggena, T. J.H.M., & Veldkampc, B. P. (2017). Formative use of test results: A user's perspective. *Studies in Educational Evaluation*, 52, 12–23.
- Lam, R. (2013). Formative Use of Summative Tests: Using Test Preparation to Promote Performance and Self-Regulation. *Asia-Pacific Education Researcher*, 22, 69–78.
- Machado, E. A. (2020). *Participação dos alunos nos processos de avaliação*. Folha de apoio à formação. Projeto MAIA. Instituto de Educação da Universidade de Lisboa e Direção-Geral de Educação do Ministério da Educação.
- Merle, P. (2018). *Les pratiques d'évaluation scolaire. Historique, difficultés, perspectives*. Presses Universitaires de France.
- Rawson, K. A., & Dunlosky, J. (2012). When is practice testing most effective for improving the durability and efficiency of student learning? *Educational Psychology Review*, 24, 419–435.



Cofinanciado por: