

# Avaliações externas e aprendizagens dos alunos: uma reflexão crítica

**Domingos Fernandes**

Universidade de Lisboa - Portugal

## Resumo

O propósito deste texto foi produzir uma reflexão crítica acerca das avaliações externas e das suas relações com as avaliações internas e com as aprendizagens dos alunos. Discutiram-se efeitos positivos e negativos das avaliações externas sobre as aprendizagens e o ensino e a necessidade de aprofundar e melhorar as avaliações internas. A reflexão produzida mostrou a possibilidade de se evoluir para outra geração de avaliações externas, baseada em princípios socialmente construídos e centrados nas aprendizagens dos alunos, devidamente articulada com outras avaliações (internas e externas) e enquadrada por políticas públicas materializadas em programas que apoiem a formação dos professores.

**Palavras-chave:** Avaliação em Educação. Avaliação das Aprendizagens. Avaliação Externa. Avaliação Interna. Políticas Públicas de Educação.

## Abstract

The purpose of this paper was to produce a critical reflection on external assessments and on its relationships with both internal assessments and student learning. A discussion has been made about both positive and negative effects of external assessments on learning and teaching as well as on the need to deepen and to improve internal assessments. As a result of the discussion, a reflection was made on the possibility to evolve towards another generation of external assessments that ought to be based upon social built principles focused on student learning. Besides, it should be strongly connected with other assessments (internal and external) and framed by public policies and programs that support inservice teacher training.

**Key words:** Educational Evaluation. Learning Assessment. External Assessment. Internal Assessment. Public Policies on Education.

## Introdução e enquadramento

As avaliações externas das aprendizagens dos alunos continuam a desempenhar um papel mais ou menos significativo nas políticas educativas de quase todos os países do mundo. Apesar de se poder questionar fundamentadamente o seu real valor pedagógico, a verdade é que os decisores políticos continuam a insistir na sua utilização por razões que, muitas vezes, estão associadas à ideia de que elas constituem uma medida credível da qualidade do ensino, da qualidade das aprendizagens e, em geral, da qualidade da educação. Na verdade, a convicção de que a utilização das avaliações externas contribui para reformar e melhorar a qualidade dos sistemas educativos está muito presente, mesmo bastante enraizada, em largos setores das sociedades e, talvez por isso, é politicamente poderosa. Aceita-se como conhecimento adquirido que as avaliações externas são sinônimo de rigor, de exigência e de qualidade sem equacionar a possibilidade da sua utilização poder ter uma diversidade de efeitos nefastos.

Por outro lado, para uma diversidade de autores e intervenientes sociais e políticos, as avaliações externas são percebidas muito negativamente, argumentando-se que aprofundam as desigualdades entre os alunos e entre as escolas. Não se equaciona sequer a possibilidade de se poderem desenvolver avaliações externas cujos resultados possam ser utilizados para apoiar os esforços de professores e alunos para melhorar o ensino e as aprendizagens.

Temos assim duas posições polarizadas que parecem estar mais dependentes das crenças, concepções, valores e ideologias dos seus defensores do que do conhecimento que, apesar de tudo, vem sendo produzido no domínio das avaliações externas e das suas relações com uma grande diversidade de variáveis (por ex. aprendizagens, práticas de ensino, práticas de avaliação). Tem de reconhecer-se, obviamente, que a decisão de utilizar avaliações externas está indelevelmente associada às opções políticas e ideológicas de quem tem a responsabilidade de as pôr em prática. Mas não se podem ignorar desenvolvimentos que, nos últimos anos, se têm verificado quer na natureza das provas, quer na utilização que se faz dos seus resultados, quer ainda na natureza das suas relações com as avaliações internas.

Antes de prosseguir, interessa clarificar que, no contexto deste artigo, entende-se que a avaliação externa é aquela que é de responsabilidade de uma entidade qualquer externa à escola (Instituto de Avaliação, Ministério da Educação, Agência Independente de Avaliação) e tem como principal propósito medir, num dado momento, o que os alunos sabem e são capazes de fazer em um ou mais domínios do currículo de um ou mais anos de escolaridade. Assim, todos os processos de concepção, elaboração, distribuição, administração, coleta, correção e divulgação dos resultados são de integral responsabilidade de uma instituição externa às escolas. É igualmente importante esclarecer que os resultados das avaliações externas das aprendizagens podem ser utilizados de formas muito diferentes. Em muitos países eles podem, por exemplo, ser utilizados para efeitos da progressão acadêmica dos alunos, chegando

a ter um “peso” de mais de 80% na classificação final, para efeitos de avaliação das escolas e/ou de avaliação dos professores. Noutros casos são utilizados para efeitos da progressão académica dos alunos, mas com um “peso” muito reduzido na sua classificação final, podendo também ser utilizados para efeitos de avaliação das escolas e dos professores mas com ponderações consideradas baixas. Os resultados das avaliações externas podem ainda ser utilizados com o principal propósito de contribuir para avaliar o desempenho global do sistema educativo e/ou fazer um ponto de situação e proporcionar *feedback* aos alunos, às famílias, aos professores e às escolas acerca das aprendizagens dos alunos. Neste caso, os resultados não são utilizados para efeitos de progressão académica dos alunos e/ou da avaliação dos professores e das escolas.

Assim, de modo geral, conforme o grau com que os seus resultados são utilizados na progressão académica dos alunos ou na avaliação das escolas e/ou professores, as avaliações externas podem ser consideradas como sendo de elevado, moderado ou reduzido impacto (em língua inglesa utilizam-se, respetivamente, as expressões *high stakes*, *moderate stakes* e *low stakes*). Por exemplo, em Portugal, os exames nacionais de matemática e de língua portuguesa do final do 9.º ano de escolaridade, que têm uma ponderação de 30% para efeitos da classificação final em cada uma destas disciplinas, são normalmente considerados de reduzido impacto (*low stakes assessment*). As chamadas provas de aferição realizadas nos 2.º, 5.º e 8.º anos de escolaridade numa diversidade de disciplinas, que vão sendo alteradas anualmente e cujos resultados individuais são devolvidos a cada aluno e os resultados globais das escolas a cada escola, são também consideradas de reduzido ou mesmo muito reduzido impacto. Já os exames do ensino secundário, realizados nos 11.º e 12.º anos de escolaridade e com uma ponderação de 30% para efeitos da certificação da conclusão daquele ciclo de estudos, mas com uma ponderação de 50% para efeitos do cálculo da classificação de acesso ao ensino superior, são normalmente considerados de impacto moderado (*moderate stakes assessment*).

Para efeitos deste trabalho interessa referir que a avaliação do que os alunos sabem e são capazes de fazer é, em geral, orientada por propósitos tais como: a) contribuir para melhorar as aprendizagens e o ensino; b) certificar as aprendizagens e competências desenvolvidas ao longo de um determinado período de tempo; c) recolher informação que permita conhecer o desempenho dos sistemas educativos; e d) contribuir para melhorar a qualidade da educação. Para concretizar estes e outros propósitos, normalmente relacionados com a avaliação do desempenho docente e com a avaliação das escolas, são normalmente utilizadas avaliações internas, avaliações externas ou uma qualquer articulação entre as duas.

Referindo-se ao caso de Portugal, Santiago *et al.* (2012) consideraram existir um certo equilíbrio no que se refere à avaliação das aprendizagens dos alunos decorrente da predominância da avaliação interna em relação à avaliação externa. Isto significa que o “peso” da avaliação interna e das respectivas classificações é sempre bastante superior ao “peso” da avaliação externa. Ou seja, a avaliação pedagógica, que decorre

das interações e dinâmicas que se estabelecem entre professores e alunos nas salas de aula sobrepõe-se claramente às provas de avaliação externa (por ex. exames, provas aferidas). Trata-se de uma constatação que parece fazer todo o sentido já que, à partida, são os professores que estão em melhores condições para se pronunciarem acerca do que os alunos sabem e são capazes de fazer em cada momento. Além disso, são igualmente os professores que, em princípio, poderão desenvolver uma avaliação tendencialmente contínua, fortemente articulada com o ensino e com a aprendizagem e que permita a distribuição de *feedback* de elevada qualidade a todos e a cada um dos alunos para apoiar os seus esforços de aprendizagem. Isto é, as avaliações internas de natureza eminentemente formativa, de integral responsabilidade dos professores, poderão ser avaliações orientadas para as aprendizagens dos alunos, para a melhoria da qualidade dessas mesmas aprendizagens. Consequentemente, pode avaliar-se para melhorar e para ajudar a aprender e não para classificar. Na verdade, avaliar não é classificar e, neste sentido, não é medir. Ainda que, naturalmente, a avaliação nos permita recolher informação de qualidade para classificar os alunos. Mas esse não é, como se viu, o seu mais fundamental propósito.

Porém, a decisão política de instituir em um sistema escolar a predominância da avaliação interna e, em particular, da avaliação formativa ou da avaliação para as aprendizagens, não é, por si só, uma garantia de que os alunos aprendem melhor, com mais profundidade. O caso de Portugal é, a esse propósito, paradigmático pois a predominância da avaliação formativa interna está clara e formalmente instituída há mais de 20 anos no sistema escolar e o número de reprovações, apesar de ter vindo a decrescer ao longo destes anos, continua a ser um dos mais elevados da Europa, situação que persiste desde os anos 60 do século passado (Fernandes, 2005, 2007, 2009, 2014). Não será, obviamente, qualquer avaliação interna que contribui para melhorar a qualidade das aprendizagens nos sistemas escolares. Terá de ser uma avaliação com características como as que foram referidas acima em que, por exemplo, a distribuição de *feedback* e a proximidade entre professores e alunos constituem elementos incontornáveis. Avaliar com o propósito de melhorar tem profundas implicações pedagógicas que obrigam a reconfigurar de forma mais ou menos profunda aspectos tais como: a) os papéis de professores e alunos; b) as dinâmicas de trabalho nas salas de aula; c) as práticas de ensino; d) as tarefas propostas aos alunos; e e) a participação dos alunos nos processos de ensino, avaliação e aprendizagem. As exigências deste tipo de avaliação têm sido amplamente discutidas na literatura (veja-se em Black & Wiliam, 2006a; Fernandes, 2008; Gipps, 1994; Luckesi, 2006; Marinho, Leite e Fernandes, 2013, 2014; Vasconcellos, 2006). O que a discussão decorrente da pesquisa na área nos tem mostrado é que os professores e as escolas têm revelado uma diversidade de dificuldades para mudar as suas práticas de avaliação no sentido de torná-las um meio privilegiado para que os alunos possam aprender mais e melhor. Para além disso, a pesquisa também tem identificado outras dificuldades relacionadas com as avaliações internas, nomeadamente no que concerne aos seus níveis de consistência e de validade (Stobart, 2006, 2008). Parece assim ser relevante que a credibilidade das avaliações internas possa ser objeto de discussão e de reflexão tendo em conta o seu real potencial para melhorar muito

significativamente as aprendizagens dos alunos e a qualidade dos sistemas escolares.

As avaliações externas, vulgo exames ou testes que interferem na progressão acadêmica dos alunos, também não estão isentas de problemas e dificuldades de diversa natureza (técnica e tecnológica, pedagógica, política e social). Apesar dos reconhecidos desenvolvimentos na “tecnologia” e na “técnica” para produzir instrumentos de avaliação externa e para analisar os seus resultados, parece reunir um significativo consenso a ideia de que os efeitos nefastos provocados por este tipo de avaliações superam significativamente os seus efeitos positivos ou as suas reconhecidas vantagens (Kellaghan e Madaus, 2003; Kornhaber, 2004; Madaus, Russell e Higgins, 2009; Shepard, 1988, 2000; Stecher, 2002).

Foi tendo em conta as ideias que muito sucintamente acabaram de se discutir que se decidiu produzir este texto cujo principal propósito é produzir uma reflexão crítica que, de algum modo, contribua para desenvolver uma discussão em que se possam considerar possíveis evoluções relativas à natureza e aos propósitos das avaliações externas e às suas relações com as avaliações internas e com as aprendizagens dos alunos.

## **Acerca das avaliações externas e dos seus efeitos**

A ideia que normalmente está na base da utilização de exames, testes nacionais ou outras formas de avaliação externa é a da melhoria da qualidade da educação. Na verdade, muitos decisores políticos e setores mais ou menos alargados das sociedades acreditam que a definição e a medição de determinado tipo de metas ou *standards* de aprendizagem estão associadas à melhoria do ensino e das aprendizagens. Dito de outro modo, crêem que quando há um sistema de avaliação externa baseado em exames ou em outro tipo de instrumentos, destinado a medir o desempenho dos alunos relativamente a certas metas preestabelecidas, os professores e os alunos tendem, respetivamente, a ensinar e a aprender melhor. Esta é uma crença que, um pouco por todo o mundo, talvez tenha atingido o auge da sua popularidade nos anos 70 e 80 do século passado e que, no fundo, consistia na ideia de que as avaliações externas eram um elemento-chave para concretizar as reformas nos sistemas escolares. A verdade é que as evidências acabaram por comprovar que as reformas não viram concretizados os seus mais fundamentais propósitos como, por exemplo, desenvolver as capacidades superiores de pensamento dos alunos ou as suas competências para resolver uma diversidade de problemas. As reflexões de Lauren Resnick e de outros investigadores a este propósito permitem inferir que as avaliações externas que foram sendo utilizadas nem foram capazes de avaliar uma diversidade de competências superiores de pensamento dos alunos nem conseguiram contribuir de forma clara para concretizar os propósitos mais fundamentais das reformas (Kornhaber, 2004; Resnick, 1987; Stecher, 2002)

O chamado efeito de *backwash* ou de *washback*, consequências mais ou menos negativas que resultam da utilização das avaliações externas, tem sido objeto de investigação e discussão, particularmente nas últimas três décadas. As conclusões de muitas das pesquisas realizadas (Shepard, 2000; Stecher, 2002) permitiram identificar e caracterizar uma diversidade de efeitos nefastos dos exames, alguns dos quais se discutem em seguida.

Um dos efeitos mais nefastos das avaliações externas é o chamado “estreitamento” ou “afunilamento” do currículo que consiste na tendência de os professores ensinarem para o que consideram que vai ser perguntado nos exames. Nestas condições, todo o resto é ignorado ou tratado de forma superficial. Além do mais, este efeito afeta substancialmente o que é ensinado e como é ensinado, dando, por exemplo, mais ênfase a tarefas de natureza rotineira e menos a tarefas que exigem mais tempo e que envolvem processos de recolha, organização, tratamento e análise de informação. Nestes termos, as disciplinas que não são objeto de exame tendem a perder a sua importância aos olhos de alunos, professores, pais e encarregados da educação, havendo assim um certo “esvaziamento” do currículo. Ou seja, domínios que são relevantes para a formação dos alunos são tratados superficialmente ou podem nem sequer ser ensinados ao longo da sua escolaridade, pois só se atribui valor e, por isso, só se ensina o que é examinado (Abu-Alhija, 2007; Casassus, 2009; Fullan, 2009). A investigação também sugere que a tendência para “estreitar” o currículo é tanto maior quanto mais consequências forem atribuídas aos resultados dos exames. Por exemplo, se os exames têm um elevado “peso” para efeitos da progressão académica dos alunos, então a tendência para ensinar e avaliar para o exame será muito maior do que no caso em que esse mesmo exame tem um “peso” moderado ou fraco. Além disso, a pesquisa tem também mostrado que as escolas inseridas em comunidades mais deprimidas dos pontos de vista social, económico e cultural tendem a utilizar mais tempo a “ensinar para o teste” e, por isso, o empobrecimento do currículo é superior ao das escolas inseridas em meios mais favorecidos.

As avaliações externas condicionam de forma mais ou menos significativa a seleção das propostas de trabalho que se propõem aos alunos, levando muitos professores a utilizarem frequentemente tarefas com questões semelhantes às que normalmente aparecem nos exames. Ou seja, nas investigações realizadas, a maioria dos professores respondem invariavelmente que, ao longo do ano, de forma sistemática e regular, se preocupam em preparar os alunos para responderem às questões dos exames. O treino de conteúdos e capacidades específicas através da repetição de exercícios é o que predomina nas aulas. As tarefas que exigem tempo e que suscitam a mobilização, a integração e a utilização de conhecimentos assim como a sua discussão são normalmente relegadas para segundo plano ou mesmo ignoradas. Os resultados das avaliações externas aos quais os meios de comunicação social e/ou as pressões de natureza política atribuem o “estatuto” de *high stakes* não traduzem em geral o que os alunos sabem e são capazes de fazer. Na verdade, os alunos são treinados para responder a questões que supostamente são formuladas nos exames e, conseqüentemente, tendem a obter respostas certas. No entanto,

tal como referem vários pesquisadores, responder acertadamente às questões dos exames não significa necessariamente que haja reais aprendizagens (Fernandes, 2015; Gipps, 1994; Shepard, 1988, 2000).

Um outro efeito bem identificado e bastante preocupante tem a ver com a tendência para as escolas (públicas e privadas) utilizarem estratégias mais ou menos explícitas no sentido de “afastar” os alunos ditos com “mais dificuldades”, isto é, os alunos que, supostamente, terão poucas possibilidades de ter sucesso nos exames. Isto sucede sobretudo quando as avaliações externas são percebidas como *high stakes* e/ou os resultados são utilizados para produzir *rankings* das escolas. Então, as escolas dificultam ou impedem as matrículas a alunos com percursos acadêmicos menos favoráveis ou utilizam uma diversidade de processos para que os alunos que frequentam a escola desistam ou cancelem as suas matrículas, apresentando-se a exame a título individual. (Em Portugal estes alunos são designados como “autopropostos” uma vez que se inscrevem nos exames por sua iniciativa e sem qualquer enquadramento institucional.) Quase seria escusado referir que a esmagadora maioria desses alunos são provenientes de famílias de baixo estatuto social e econômico. São práticas que resultam das visões e das pressões mercantis que as sociedades têm vindo a desenvolver em relação à educação. Em suma, os alunos que merecem maior atenção e apoio por parte das escolas são, em geral, aqueles que, supostamente, dão mais garantias de poder ter sucesso nos exames. Os que não dão essas garantias ficam, num certo sentido, entregues a si próprios, acabando por cancelar as suas matrículas ou por reprovar. Nestas condições, as avaliações externas, particularmente os exames, podem contribuir de forma mais ou menos significativa para o desenvolvimento de processos que excluem e discriminam os alunos, particularmente os que provêm de meios social e economicamente mais deprimidos.

Finalmente, é importante referir que os próprios professores sofrem consequências como resultado das pressões sentidas para que os seus alunos obtenham boas notas nos exames. Um número de professores considera que os exames acabam por facilitar as suas tarefas, pois permitem-lhes identificar o que é importante ensinar e, assim, focarem-se nos conteúdos curriculares que são objeto de exame. Trata-se de “ensinar para o exame” que é, reconhecidamente, uma visão redutora e simplista do que deve ser a educação e a formação dos alunos. Neste sentido, ao concentrarem os seus esforços no que “cai no exame” e na obtenção de bons resultados, os professores poderão não ter condições para criar reais oportunidades para que todos e cada um dos seus alunos tenham acesso a aprendizagens mais abrangentes e significativas. O foco é nas classificações e não na qualidade das aprendizagens e, assim sendo, as capacidades e competências pedagógicas dos professores acabam por se desvalorizar e degradar. Além disso, os *rankings* produzidos através das classificações geram comparações entre escolas e professores que podem igualmente ter efeitos nefastos nas relações entre instituições e entre profissionais (Madaus, Russell e Higgins, 2009).

É ainda importante referir que as avaliações externas não estão diretamente relacionadas com a melhoria das aprendizagens dos alunos. Na verdade, os seus

principais propósitos têm fundamentalmente a ver com ações tais como: a) verificar ou controlar se os conteúdos previstos no currículo são ensinados e aprendidos pelos alunos; b) utilizar os resultados obtidos pelos alunos como forma de “pedir contas” e/ou de responsabilizar os professores e as escolas; c) certificar as aprendizagens e/ou competências demonstradas pelos alunos; e d) selecionar alunos para efeitos de progressão académica para outro nível de ensino como é, em Portugal, o caso no acesso ao ensino superior (Fernandes, 2007; Kellaghan e Madaus, 2003).

As avaliações externas e, em particular, os exames, possuem algumas características que têm sido vistas como positivas ou como vantagens por parte de um número de autores (Kellaghan e Madaus, 2003; Madaus, Russell e Hastings, 2009; Stobart, 2008). Tendo em conta que a qualidade das classificações internas é pouco conhecida e havendo uma diversidade de razões para considerar que pode estar aquém do desejável, os resultados das avaliações externas podem ter um efeito moderador, contribuindo para melhorar a validade e a consistência das classificações realizadas no âmbito das escolas pelos seus professores. Porém, este aspecto positivo deve ser analisado com algum cuidado uma vez que a fiabilidade (consistência) dos resultados das avaliações externas é pouco investigada e largamente desconhecida. Pode especular-se que talvez se receie que os resultados de uma pesquisa acerca da consistência dos resultados dos exames, por exemplo, poderia desacreditá-los e descredibilizá-los perante largos setores da sociedade. Mas a verdade é que os poucos estudos que se conhecem apontam no sentido de considerar que a fiabilidade dos resultados tende a ser mais baixa do que se poderia supor ou seria desejável. A validade das avaliações externas está mais estudada e parece ser mais credível, sobretudo a validade de conteúdo. Por outro lado, no caso das avaliações internas formativas, a validade, que deve ser pensada em termos das suas consequências, terá a ver com a melhoria das aprendizagens dos alunos. Digamos que a avaliação formativa interna será válida se, em consequência dela, houver realmente aprendizagem. A fiabilidade das avaliações formativas internas, ainda que seja relevante, não assume a importância nem tem efeitos semelhantes ao que sucede nas avaliações externas. Na verdade, no contexto de sala de aula, os professores podem sempre ir “ajustando” os seus juízos acerca do que os alunos sabem e são capazes de fazer ao longo do processo de ensino, aprendizagem e avaliação. Estas questões têm sido debatidas em alguma literatura, mas carecem, claramente, de aprofundamento e de clarificação (Black e Wiliam, 2006b; Stobart, 2006). Em todo o caso, sabemos bem mais acerca da fiabilidade e da validade dos resultados das avaliações externas do que dos resultados das avaliações internas.

Em suma, os conceitos de validade e de fiabilidade são incontornáveis para garantir a qualidade das avaliações em geral, mas são particularmente relevantes no contexto das avaliações externas consideradas *high stakes*, tendo em conta as suas consequências. Os resultados das avaliações externas não podem, obviamente, ser revertidos e os seus efeitos podem ser muito nefastos para a vida dos alunos, das escolas ou dos professores (Black, 1998).

Para uma diversidade de autores (Resnick, 1987; Wiggins, 2011), os testes ou exames nacionais podem ter um relevante papel na indução de práticas inovadoras de ensino e de avaliação. Por exemplo, a “avaliação autêntica” proposta por Wiggins (2011), se utilizada em larga escala, poderia proporcionar oportunidades para que os professores, ao “ensinarem para o teste”, estivessem na verdade a trabalhar com os seus alunos capacidades complexas de pensamento. Estaríamos assim perante uma “educação para aprender a pensar” que poderia ser induzida pela natureza dos testes ou dos exames e dos itens que os integram, como nos sugeriu um dia Lauren Resnick (Resnick, 1987). São igualmente conhecidos efeitos positivos resultantes da administração de provas aferidas (*low stakes*) e de exames (*moderate stakes*) ao nível das práticas de ensino dos professores (Poinha, 2012; Stobart, 2008).

Há um número de autores que referem à importância das avaliações externas, nomeadamente do tipo *low stakes*, na recolha de informação acerca dos desempenhos dos sistemas educativos para melhorar os processos de tomada de decisão. Esta ideia está naturalmente relacionada com a necessidade sentida pelos políticos e decisores em geral de que é importante recolher informação acerca das aprendizagens dos alunos num dado momento para que se possa conhecer e compreender “o que acontece” e, conseqüentemente, poder tomar decisões mais adequadas e fundamentadas (Madaus, Russell e Stobart, 2008).

As avaliações externas podem também constituir um meio relevante para proporcionar *feedback* de qualidade aos alunos, professores, gestores escolares, pais e encarregados de educação. Na verdade, podem contribuir para que todos se mobilizem para, respectivamente, alterar e/ou reorientar os seus esforços de estudo, de ensino, de organização e funcionamento pedagógico das escolas e de apoio aos seus filhos e educandos (Madaus, Russell e Higgins, 2009; Poinha, 2012). Em particular, certo tipo de avaliações externas *low stakes*, ou com nenhum ou poucos efeitos na progressão académica dos alunos ou na avaliação das escolas e professores, podem realmente ser úteis aos principais intervenientes, como é o caso das atuais provas aferidas no sistema educativo português (Fernandes, 2015).

Para concluir esta secção em que se procurou caracterizar fundamentadamente a avaliação externa através da sua natureza, dos seus propósitos e dos seus efeitos, parece fazer sentido referir que, de acordo com Madaus, Russell e Higgins (2009), os exames e as avaliações externas em geral constituem um sistema paradoxal. Na verdade, por um lado, podem ser concebidas para produzir efeitos positivos que são deliberadamente assumidos e desejados tais como: a) induzirem práticas inovadoras de ensino e de avaliação; b) contribuir para que os conteúdos constantes no currículo sejam efetivamente lecionados; c) clarificarem junto dos professores, das escolas e dos alunos as questões que são consideradas essenciais e fundamentais em cada um dos domínios do currículo; e d) incentivarem os alunos a estudar e a desenvolverem os esforços necessários para aprenderem os conhecimentos e capacidades que estão previstas no currículo. Mas, por outro lado, as avaliações externas produzem efeitos, geralmente bastante negativos, que não foram previstos, nem desejados, nem

deliberados e que normalmente prejudicam seriamente a qualidade da educação e da formação dos alunos tais como: a) o “estreitamento” do currículo que decorre da tendência para ensinar apenas o que, supostamente, é objeto dos exames; b) o abandono escolar precoce; c) a utilização de estratégias indesejáveis por parte das escolas para afastar os alunos que não deem garantias de ter sucesso nos exames; e d) a focagem nas classificações em vez de nas aprendizagens dos alunos.

Perante a natureza paradoxal das avaliações externas e sabendo-se que os seus efeitos negativos superam, em geral, os seus efeitos positivos, parece evidente que urge refletir acerca do que poderão ser os desenvolvimentos futuros quanto à necessidade real das avaliações externas, quanto à sua natureza e propósitos e à sua relação com as avaliações internas e com as aprendizagens dos alunos. É o que se fará a seguir.

## **Para uma reflexão acerca da reconfiguração das avaliações externas**

A discussão anterior ilustrou que os académicos e outros intervenientes nos sistemas escolares estão claramente divididos entre duas posições extremas no que concerne à utilização de avaliações externas, muito particularmente aquelas cujos resultados produzem efeitos elevados ou moderados (*moderate stakes* e *high stakes*). Na verdade, enquanto uns defendem os exames alegando que eles clarificam o que é preciso ensinar e aprender e, assim, ajudam professores e alunos a centrarem os seus esforços de ensino e de aprendizagem, outros argumentam que eles causam o empobrecimento do currículo porque tudo está focado nas classificações e no que é perguntado nos exames com irremediáveis prejuízos para os alunos (Gipps, 1994; Kellaghan, 2003, Kellaghan e Madaus, 2003, Kellaghan e Greaney, 2001). Ou seja, enquanto uns defendem que os exames melhoram a qualidade da educação, outros defendem precisamente o contrário. Porém, em rigor, nenhuma destas posições antagónicas foi ainda irrefutavelmente demonstrada com base em sólida evidência científica. Trata-se de uma área em que os valores, as crenças, as concepções filosóficas e as ideologias, assim como a argumentação e a persuasão dos seus protagonistas, têm uma significativa importância nas opções que acabam por ser tomadas nos sistemas escolares. Para uma discussão aprofundada sobre esta questão crítica vejam-se, por exemplo, os trabalhos de autores tais como Esteban (2009) Harlen (2007), Kellaghan e Grisay (1995), Kornhaber (2004) e Stobart (2008).

Tendo em conta estas considerações, a questão que se coloca é a de saber se é possível conceber e pôr em prática um sistema de avaliação externa qualquer em que se reduzem ao mínimo as suas desvantagens e os seus efeitos indesejáveis e nefastos, mantendo os seus efeitos positivos. Parece razoável reconhecer-se que a maioria das avaliações externas podem ser utilizadas para auditar e diagnosticar determinados aspectos dos sistemas escolares, mas, com raras exceções, muito dificilmente podem ser consideradas pedagogicamente adequadas para ensinar, avaliar e aprender melhor. As avaliações internas, que ocorrem nas salas de aula e são

de integral responsabilidade dos professores, são capazes de melhorar estes processos basilares mas, através delas, muito dificilmente se conseguem auditar os sistemas escolares. Daqui parece decorrer a necessidade de promover a sua articulação, construída com base na ideia de que as aprendizagens a desenvolver pelos alunos são a mais fundamental prioridade e de que é igualmente fundamental caminhar no sentido de eliminar os efeitos negativos das avaliações externas, sobretudo as de forte impacto (*high stakes*). Ou seja, tal como é proclamado por muitos autores, podem ser desenvolvidos esforços para articular as avaliações externas e as avaliações internas, assumindo a sua complementaridade (Kornhaber, 2004; Madaus, Russell e Higgins, 2009; Stobart, 2008). Para Gordon Stobart, por exemplo, as avaliações externas podem ter efeitos positivos na melhoria das práticas pedagógicas, sendo necessário trabalhar-se no sentido de os melhorar e não de os abandonar (Stobart, 2008).

Como já tive oportunidade de afirmar em outro texto (Fernandes, 2015), e se depreende da discussão anterior, a utilização isolada das avaliações externas ou das avaliações internas muito dificilmente poderá contribuir para melhorar a qualidade da educação e, em particular, o ensino e as aprendizagens. Ou seja, tendo em conta que nenhuma delas, por si só, pode contribuir para a resolução dos problemas que afetam os sistemas escolares, então parece que a sua articulação e a sua relação poderão ser um caminho a prosseguir. Para que este desafio possa ser concretizado é necessário, naturalmente, considerar um conjunto de princípios e ações que contribuam para um equilíbrio desejável entre as avaliações externas e internas e tendo em conta que as aprendizagens e a formação de todos e cada um dos alunos deverão estar no cerne dos esforços a desenvolver. Vejamos alguns desses princípios e ações.

Para que as avaliações externas possam contribuir para melhorar as aprendizagens dos alunos, tem de haver uma forte e inequívoca aposta política para melhorar as práticas de avaliação que ocorrem nas escolas e nas salas de aula. Isto significa uma forte aposta na formação contínua dos professores desenvolvida em modalidades baseadas em projetos de intervenção realizados nas escolas e com a participação ativa dos professores. Deste modo, fará sentido considerar que as avaliações externas podem contribuir para melhorar a qualidade dos sistemas escolares através da sua articulação com avaliações internas cujo principal propósito é melhorar o ensino e a aprendizagem. Mas, para tal, estas terão de ser mais credíveis, isto é, mais válidas e mais consistentes, através de sistemas engendrados com as escolas e nas escolas em que se trabalham os critérios a considerar a articulação de processos e estratégias de avaliação a utilizar e, em geral, as ações pedagógicas a desenvolver. A credibilização das avaliações internas é relevante para a promoção da articulação com as avaliações externas que, obviamente, também terão de ser mais credíveis, nomeadamente ao nível da sua validade pedagógica.

Tendo em conta as reconhecidas limitações das avaliações externas e internas, será sempre desejável articular inteligentemente os resultados obtidos através de cada uma delas. Trata-se de triangular a informação, princípio que contribui para melhorar a qualidade da informação de que se necessita para tomar decisões

mais fundamentadas. Além disso, de acordo com Kornhaber (2004), as decisões que impliquem a certificação dos alunos, o seu ingresso noutra nível de ensino, a sua retenção ou o seu ingresso em um programa de recuperação qualquer, nunca deveriam ser tomadas com base num único resultado proveniente de uma avaliação externa. Por isso mesmo, a articulação entre os resultados das avaliações internas e externas deverão ser aprofundadas tendo em conta que é através das avaliações internas que os alunos podem aprender com mais significado e profundidade. Esta situação já existe há alguns anos numa diversidade de países, inclusive Portugal, em que, por exemplo, os resultados das avaliações internas, para efeitos de progressão académica dos alunos e para efeitos de certificação, têm um peso de 70% contra 30% das avaliações externas. Reconhece-se, deste modo, o primado das avaliações e das classificações que são de responsabilidade dos professores e das escolas em detrimento das avaliações externas cujo “peso” é bastante reduzido (Fernandes, 2009; Gardner, 2006).

Uma outra questão crítica e sensível está relacionada com as reais oportunidades que os alunos têm para aprender. Esta questão tem a ver com a capacidade dos sistemas escolares assegurar as devidas condições para que todos os alunos tenham acesso a tais oportunidades. Nesta matéria há diferenças que podem ser muito significativas levando-nos a questionar, por exemplo, se os alunos tiveram oportunidades para aprender e não aprenderam ou se, na verdade, não tiveram essas oportunidades. São situações muito diferentes e que merecem ser analisadas e devidamente ponderadas. O princípio a seguir será, com certeza, o de que têm de ser criadas condições para que haja equidade nas oportunidades criadas para que todos os alunos possam aprender. Este princípio é muito exigente e obriga a um forte investimento das políticas públicas de educação para conceber e pôr em prática programas que possam concretizar aquele crucial objetivo. Repare-se que as avaliações externas, ao contrário do que por vezes é afirmado, inclusive por responsáveis políticos, não podem garantir nunca a equidade das oportunidades de aprendizagem. As avaliações externas não são, por definição e natureza, um meio para a resolução dos problemas dos sistemas escolares, nomeadamente os que se relacionam com as aprendizagens dos alunos. Apesar disso, há uma diversidade de intervenientes políticos, e não só, que parecem insistir na ideia de que uma única medida de política, como é a criação de uma avaliação externa qualquer do tipo *high stakes*, com efeitos, por exemplo, na progressão académica dos alunos e na carreira dos professores, pode resolver os problemas de ensino e de aprendizagem. Parece que ainda não compreenderam que o problema é bem mais complexo e que, por isso, exige uma diversidade de medidas de política mais elaboradas e fundamentadas que possam lidar com as questões subjacentes de natureza social, económica e cultural.

Em suma, tendo em conta a natureza, os propósitos e as reconhecidas limitações, quer das avaliações internas, quer das avaliações externas, desenvolveu-se uma reflexão e uma linha de raciocínio em que se reconhece vantagem na sua articulação. Isso implica, naturalmente, melhorias conducentes à credibilização das avaliações internas e externas obedecendo a princípios em que as aprendizagens dos alunos

estão no cerne dos esforços a desenvolver. Mas exige também políticas públicas de educação que apostem na formação contínua dos professores e na criação de reais oportunidades de aprendizagem para todos os alunos.

## Conclusão

Lauren Resnick disse-nos um dia que só se consegue obter aquilo que conseguimos avaliar e, conseqüentemente, não se obtém o que não se avalia. Esta asserção tem um grande significado, pois se, por exemplo, nos limitarmos a avaliar tarefas rudimentares e rotineiras, as aprendizagens resultantes serão de nível elementar. Os alunos são “preparados” apenas para repetir procedimentos e, se quisermos, resolver tarefas que requerem o mínimo de elaboração cognitiva. Nestas condições, quaisquer que sejam as avaliações, internas ou externas, é necessário compreender que elas têm de ser sistematicamente melhoradas e orientadas para o que se quer, e como se quer, que os alunos aprendam (Resnick, 1987).

O pensamento desta cognitivista norte-americana leva-nos a considerar que, através de todas as avaliações, deveremos centrar-nos na apresentação de propostas de ensino, de aprendizagem e de avaliação através das quais os alunos têm oportunidade para desenvolver as suas capacidades mais complexas de pensamento. A ideia é que a avaliação, externa ou interna, tem de estar sempre associada a práticas pedagógicas mais consentâneas com o conhecimento que tem sido produzido nesta área (Gardner, 2006; Gipps, 1994; Stobart, 2008). Assim, nunca será satisfatório termos avaliações externas que sejam irrepreensíveis dos pontos de vista técnico e tecnológico mas que não tenham em conta os desafios de uma pedagogia orientada no sentido acima proposto. De igual modo, as avaliações internas têm de ser profundamente melhoradas para não se limitarem a fazer a emulação dos exames, centrando-se nas classificações em vez de na melhoria das aprendizagens.

A discussão produzida neste texto permitiu afirmar que, ao contrário das crenças de uma diversidade de protagonistas, as avaliações externas, em si mesmas, não podem resolver os problemas que afetam os sistemas escolares, nomeadamente os que se relacionam com o ensino e a aprendizagem. Assim, defende-se a ideia de que temos de evoluir para uma outra geração de avaliações externas, baseada em princípios socialmente construídos e centrados nas aprendizagens dos alunos, devidamente articulada com outras avaliações (internas e externas) e devidamente enquadrada por políticas públicas materializadas em programas que apoiem a formação dos professores. Experiências recentes em alguns países sugerem que, pelo menos certo tipo de avaliações externas, como é o caso das provas de aferição dos 2.º, 5.º e 8.º anos de escolaridade postas em prática em 2015/2016 em Portugal indiciam que poderemos efetivamente estar a evoluir nesse sentido. Ao mesmo tempo que se recolhe informação acerca das aprendizagens realizadas em diferentes disciplinas para se conhecer e procurar compreender a situação no país como um todo, cada aluno recebe um relatório simples e individual com o feedback acerca do trabalho

realizado na prova e cada escola recebe um relatório global do desempenho de todos os alunos. A ideia é que os resultados possam ser mobilizados e integrados em processos de melhoria das aprendizagens e do ensino e, neste sentido, é-lhes dada uma utilização formativa.

Finalmente, parece importante referir que o desígnio de podermos ter melhores e mais profundas aprendizagens, com mais significado, e melhores sistemas escolares, é partilhado por uma grande diversidade de protagonistas um pouco por todo o mundo. Porém, aquele desígnio surge sempre associado à utilização de avaliações externas, muito mais preponderantes em alguns países do que noutros, que, em algumas modalidades (exames com fortes consequências - *high stakes*), têm normalmente efeitos indesejáveis como se discutiu ao longo do texto. O desafio que temos pela frente é saber se somos capazes de construir um sistema bem articulado de avaliações internas e externas que contribua inequivocamente para melhorar as aprendizagens e o ensino e a qualidade da educação das crianças e dos jovens.

## Referências

- Abu-Alhija, F. (2007). Large-scale testing: benefits and pitfalls. *Studies in Educational Evaluation*, 33, 50-68.
- Black, P. (1998). Testing: Friend or foe? *Theory and practice of assessment and testing*. London: Falmer.
- Black, P. e Wiliam, D. (2006a). Assessment for learning in the classroom. In J. Gardner (Ed.), *Assessment and learning*, 9-26. London: Sage.
- Black, P. e Wiliam, D. (2006b). The reliability of assessments. In J. Gardner (Ed.), *Assessment and learning*, 119-132. London: Sage.
- Casassus, Juan (2009). Uma nota crítica sobre a avaliação estandardizada. A perda de qualidade e a segmentação social. *Sísifo. Revista de Ciências da Educação*, 9, 71-78.
- Esteban, M. T. (2009). Provinha Brasil: desempenho escolar e discursos normativos sobre a infância. *Sísifo. Revista de Ciências da Educação*, 9, 47- 55.
- Fernandes, D. (2015). Avaliações externas e melhoria das aprendizagens dos alunos: questões críticas de uma relação (im)possível. In Conselho Nacional de Educação (Ed.), *Estado da Educação 2014* (pp. 290-303). Lisboa: Conselho Nacional de Educação.
- Fernandes, D. (2014). Avaliação das aprendizagens e políticas educativas: o difícil percurso da inclusão e da melhoria. In M. L. Rodrigues (Org.), *40 anos de políticas*

- de educação em Portugal: a construção do sistema democrático de ensino (Volume I)*, 231-268. Coimbra: Almedina.
- Fernandes, D. (2013). Avaliação em educação: Uma discussão de algumas questões críticas e desafios a enfrentar nos próximos anos. *Revista Ensaio: Avaliação e Políticas Públicas em Educação*, (78) 21, 11-34.
- Fernandes, D. (2010). Acerca da articulação de perspectivas e da construção teórica em avaliação educacional. In M. T. Esteban e A. J. Afonso (Orgs.), *Olhares e interfaces: reflexões críticas sobre a avaliação*, pp. 15-44. São Paulo: Cortez.
- Fernandes, D. (2009). *Educational assessment in Portugal*. *Assessment in Education: Principles, Policy & Practice*, (16) 2, 241-261.
- Fernandes, D. (2008). *Avaliar para aprender: fundamentos, práticas e políticas*. São Paulo: Fundação Editora da UNESP.
- Fernandes, D. (2007). A avaliação das aprendizagens no sistema educativo português. *Educação e Pesquisa*, (33) 3, 581-600.
- Fernandes, D. (2005). *Avaliação das aprendizagens: desafios às teorias, práticas e políticas*. Cacém: Texto Editores.
- Fullan, M. (2009). Large-scale reform comes of age. *Journal of Educational Change*, 10(2), 101-113.
- Gardner, J. (Ed.) (2006). *Assessment and learning*. London: Sage.
- Gipps, C. (1994). *Beyond testing: towards a theory of educational assessment*. Londres: Falmer.
- Kellaghan, T. (2003). Local, national, and international levels of system evaluation. Introduction. In T. Kellaghan e D. Stufflebeam (Eds.), *International handbook of educational evaluation*, pp. 873-882. Dordrecht: Kluwer.
- Kellaghan, T. e Greaney, V. (2001). *Using assessment to improve the quality of education*. Paris: UNESCO.
- Kellaghan, T. e Grisay, A. (1995). International comparisons of student achievement: problems and prospects. In OECD (Ed.), *Measuring what students learn*, pp. 41-61. Paris: OECD.
- Kellaghan, T. e Madaus, G. (2003). External (public) examinations. In T. Kellaghan e D. Stufflebeam (Eds.), *International handbook of educational evaluation*, 577-602. Dordrecht: Kluwer.

- Kornhaber, M. (2004). Appropriate and inappropriate forms of testing, assessment, and accountability. *Educational Policy*, 18 (1), 45-70.
- Luckesi, C. (2006). *Avaliação da aprendizagem escolar*. São Paulo: Cortez.
- Madaus, G., Russell, M. e Higgins, J. (2009). The paradoxes of high stakes testing: how they affect students and their parents, teachers, principals, schools and society. Charlotte, NC: Information Age Publishing.
- Marinho, P., Leite, C., & Fernandes, P. (2013). A avaliação da aprendizagem: um ciclo vicioso de “testinite”. *Estudos em Avaliação Educacional*, 24(55), 304-334.
- Marinho, P., Leite, C., & Fernandes, P. (2014). A avaliação da aprendizagem: da pluralidade de enunciações à dualidade de concepções. *Acta Scientiarum Education (BR)*, 36(1), 153-164. doi:10.4025/actascieduc.v36i1.21018
- Poinha, M. A. (2012). *Os exames e a melhoria das aprendizagens: o discurso de professores de língua portuguesa do 3º ciclo do ensino básico*. Dissertação de mestrado em Ciências da Educação (Avaliação em Educação). Lisboa: Instituto de Educação da Universidade de Lisboa.
- Resnick, L. B. (1987). *Education and learning to think*. Washington, DC: The National Academy Press.
- Santiago, P., Donaldson, G., Looney, A. e Nusche, D. (2012). *OECD Reviews of evaluation and assessment in education: Portugal 2012*. Paris: OECD Publishing.
- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29 (7), 4-14.
- Shepard, L. (1988). Should instruction be measurement driven: a debate. Comunicação apresentada no *Annual Meeting of the American Educational Research Association*. New Orleans.
- Stecher, B. M. (2002). Consequences of large-scale, high-stakes testing on school and classroom practice. In L. S. Hamilton, B. M. Stecher, & S. P. Klein (Eds.), *Making sense of test-based accountability in education* (pp. 72-100). Santa Monica (CA): Rand Corporation.
- Stobart, G. (2008). *Testing times: the uses and abuses of assessment*. London: Routledge.
- Stobart, G. (2006). The validity of formative assessment. In J. Gardner (Ed.), *Assessment and learning*, 133-146. London: Sage.

Vasconcellos, C. (2006). *Avaliação da aprendizagem: práticas de mudança – por uma práxis transformadora*. São Paulo: Libertad.

Wiggins, G. (2011). A true test: toward more authentic and equitable assessment. *Phi Delta Kappan*, 92(7), 81-93

## **Biografia**

### **Domingos Fernandes**

Instituto de Educação, Universidade de Lisboa, Professor Titular; Coordenador de Programas de Pós-Graduação e Pesquisa em Avaliação Educacional. Avaliação Educacional.

**ORCID:** <https://orcid.org/0000-0002-3713-6484>