

Avaliações Externas e Melhoria das Aprendizagens dos Alunos: Questões Críticas de uma Relação (Im)Possível

Domingos Fernandes
Universidade de Lisboa
Instituto de Educação

Introdução

Os esforços de investigação empreendidos nas últimas décadas no domínio das avaliações externas das aprendizagens dos alunos, mostram que é difícil estabelecer relações de causa e efeito, ou mesmo estabelecer correlações significativas, entre os sistemas de exames e a melhoria da qualidade dos sistemas educativos. Apesar disso, na generalidade dos países do mundo e por uma diversidade de razões (e.g., políticas, ideológicas, culturais, simbólicas), as avaliações externas sempre desempenharam e continuam a desempenhar um papel significativo nas políticas públicas de educação. Talvez porque a *teoria* que fundamenta a sua utilização para reformar e melhorar os sistemas educativos seja persuasiva, convincente e politicamente poderosa. E isto porque, em geral, há uma predisposição quase natural por parte das sociedades para associarem as avaliações externas à exigência, ao rigor e, em última análise, à garantia de qualidade dos sistemas de ensino. Talvez porque, numa diversidade de sectores das sociedades, inclusive no académico, se considere que os sistemas de avaliação externa apresentam vantagens que podem compensar as suas reconhecidas desvantagens. Mas também porque tais sistemas são mais baratos do que os que investem deliberada e abertamente em programas sistemáticos de apoio e de formação às escolas e aos professores.

Seria ingénuo considerar-se que, em si mesmas, as avaliações externas são boas ou são más. De facto, a utilização que delas se faz, assim como os seus propósitos e as suas relações com as avaliações internas e com o currículo, é que, em muito boa medida, poderão determinar o seu real significado e o seu real impacto, nomeadamente no que se refere à melhoria dos sistemas educativos. Ou seja, no que se refere às suas relações com a melhoria da qualidade do ensino dos professores e das aprendizagens das crianças e dos jovens.

Seria igualmente ingénuo pensar-se que as avaliações internas, da responsabilidade exclusiva dos professores e das escolas, são, por si sós, a garantia da qualidade que se almeja para os sistemas educativos. Portugal é, a este propósito, um excelente exemplo de que as avaliações internas, por si sós, não são a panaceia que permite melhorar a educação e a formação dos estudantes. Na verdade, durante largos anos a avaliação interna predominou (e ainda predomina) fortemente no nosso sistema educativo, chegando a ser exclusiva em todo o ensino básico, e nem por isso deixámos de ter taxas de retenção sem qualquer paralelo no contexto europeu (Fernandes, 2007,2009). Ou seja, supostamente, os alunos não aprendiam o que eram supostos aprender. Note-se que ainda hoje, apesar da introdução de exames nacionais a Matemática e a Língua Portuguesa nos anos terminais de cada um dos três ciclos do ensino básico, a avaliação interna ainda é claramente predominante no nosso sistema de ensino. Como é óbvio, não é qualquer avaliação interna que permite apoiar os alunos nos seus processos de aprendizagem e a investigação tem mostrado dificuldades das escolas e dos professores para desenvolverem práticas de avaliação consistentes com aquele propósito fundamental; a ação política, por seu lado, não tem em geral, sido orientada para ajudar a superar tais dificuldades. Além do mais, são igualmente conhecidos outros problemas relacionados com as avaliações internas como é o caso da sua consistência, da sua validade e da sua própria credibilidade (e.g., Black e Wiliam, 2006; Stobart, 2006). No que se refere às avaliações externas, como se verá mais adiante, há também problemas bem identificados que podem questionar a credibilidade que, em geral, lhe é atribuída pela sociedade.

A ideia de que podemos ter melhores sistemas educativos, isto é, onde todos os estudantes possam aprender melhor e com mais significado, continua a orientar os esforços de investigadores, professores e outros profissionais da educação e da formação. Mas também os esforços de agentes políticos, de governantes e de uma grande diversidade de organizações/instituições e de cidadãos indiferenciados. E essa ideia traz normalmente associada a existência de sistemas de avaliação externa, mais ou menos preponderantes, mais ou menos articulados com a avaliação interna, mais ou menos decisivos para o progresso académico dos estudantes.

A questão que se poderá formular, apesar de todas as reconhecidas dificuldades, é a de saber se é possível termos sistemas de avaliação externa que contribuam para que todos os alunos possam aprender mais e, sobretudo, melhor, com mais significado.

Associada a esta questão surgem normalmente muitas outras tais como a de saber se é possível termos sistemas de avaliação externa que não excluam certos grupos, com determinadas características sociais e económicas ou outras e se as avaliações externas podem integrar sistemas mais inteligentes que as articulem melhor com as avaliações internas.

Foi a partir das considerações acima explicitadas que organizei este trabalho, explorando teoricamente a possibilidade da complementaridade das avaliações internas e externas e produzindo reflexões e recomendações para que estas últimas possam ser mais adequadas e mais relacionadas com a melhoria das aprendizagens dos estudantes e da qualidade dos sistemas educativos.

Perspetivas Sobre O Domínio Do Conhecimento Da Avaliação

A avaliação em educação é um domínio do conhecimento que se tem vindo a estabelecer e a consolidar na academia sobretudo a partir de meados do século passado, na sequência do *vendaval* que assolou o chamado mundo ocidental, após o lançamento do *sputnick* por parte da então União Soviética. Nessa altura, todos os sistemas educativos ocidentais iniciaram profundas reformas, sobretudo nas ciências físicas e naturais e na matemática, destinadas a melhorar a competência tecnológica do *ocidente*. O desenvolvimento dessas reformas e dos programas de ação a elas associados, foi invariavelmente monitorizado através da utilização sistemática de estudos de avaliação. Consequentemente, a avaliação como disciplina desenvolveu-se significativamente tendo-se estabelecido os seus conceitos estruturantes como é o caso da avaliação formativa e da avaliação sumativa, introduzidos em 1967 por Michael Scriven no contexto da avaliação de programas e, em 1971, em contextos de avaliação das aprendizagens, por Benjamim Bloom e colaboradores (Bloom Hastings e Madaus, 1971; Scriven, 1967). Por outro lado, foram sendo discutidos e clarificados os seus propósitos (e.g., discernir qualidade, compreender, certificar, melhorar, comparar, ajuizar), as suas funções (e.g., informar a sociedade acerca da qualidade dos objetos avaliados, orientar tomadas de decisão), o papel dos avaliadores e dos participantes nos processos de avaliação (e.g., mais ou menos intervenientes, mais ou menos distanciados, mais ou menos empenhados, mais ou menos comprometidos com causas sociais e/ou políticas).

Qualquer discussão acerca da avaliação como disciplina, como domínio do conhecimento, não pode ignorar as questões epistemológicas, ontológicas e metodológicas que lhe são subjacentes e que, no fundo, determinam a existência de uma diversidade de perspectivas e de abordagens como, por exemplo, as que foram definidas por Stufflebeam (2000). A construção teórica no domínio da avaliação reside precisamente no desenvolvimento de tais abordagens de avaliação a partir das práticas e das perspectivas ideológicas, filosóficas e sociais dos respectivos autores (Fernandes, 2010, 2013). De modo geral, as avaliações podem ser mais ou menos inspiradas na lógica da ciência ou em lógicas emergentes associadas a racionalidades de natureza crítica. A lógica clássica da avaliação, que segue a lógica da produção científica de conhecimento, é bastante simples e desenvolve-se em quatro etapas: a) definição de critérios; b) estabelecimento de standards ou normas de consecução dos critérios; c) definição de um processo de medida; e d) formulação de um juízo acerca do valor e do mérito do objeto avaliado através da análise feita aos resultados obtidos na etapa anterior. Trata-se de uma lógica baseada em critérios em que a medida e os esforços para se conseguirem obter avaliações tão objetivas quanto possível assumem uma particular relevância. Assim, nas avaliações baseadas no pensamento criterial predominam os chamados métodos quantitativos e, por natureza, a participação dos diversos intervenientes e, em particular, dos avaliados é, em regra, bastante limitada. Além do mais, estão mais centradas nos resultados e os avaliadores procuram ser neutros, isto é, procuram manter-se distantes em relação às realidades ou aos fenómenos sociais que estão a estudar para que, desse modo, não os *contaminem* nem se deixem *contaminar* por eles. Muitas das avaliações contemporâneas, como é o caso das avaliações externas, assentam significativamente nesta lógica criterial de avaliação que, como se imagina, exige um grande esforço na definição de critérios, que é uma tarefa bastante complexa e difícil, e na construção de instrumentos que permitam, digamos assim, medir a consecução dos critérios tendo em conta os standards ou normas estabelecidos. Esta última tarefa, no contexto da avaliação externa das aprendizagens, é igualmente de grande complexidade e, como é sabido, tem-se baseado na Teoria Clássica dos Testes e, mais recentemente, na Teoria de Resposta ao Item. Medir rigorosamente o que os estudantes sabem e são capazes de fazer exige a mobilização de uma diversidade de disciplinas (e.g., pedagogia, currículo, avaliação, psicologia, estatística) e, ainda assim, os seres humanos não podem afirmar, acima de qualquer suspeita, que um dado teste mede exatamente o

que os alunos sabem ou são capazes de fazer relativamente a um dado domínio do currículo (ver, por exemplo, Madaus, Russell e Higgins, 2009). Isto não significa, obviamente, que os testes ou os exames não produzam resultados credíveis, plausíveis, justos e úteis. O que significa é que os testes ou os exames, se quisermos, não são à prova de imprecisões e até de erros, não podemos dizer que são um valor absoluto e isso tem que ser tido na devida conta. Por outro lado, o tratamento estatístico que se faz aos resultados das avaliações em larga escala, incluindo as de natureza internacional, também não está isento de problemas que necessitam de ser resolvidos (ver, por exemplo, Goldstein, 1996, 2004).

Nos últimos anos, face às limitações das abordagens criteriosais, pois nem tudo pode ser objeto de medida, e ao crescente reconhecimento e aceitação de epistemologias e de racionalidades de raiz mais crítica e sociocrítica na construção do conhecimento, passaram a valorizar-se abordagens de avaliação baseadas nas experiências pessoais das pessoas, nas suas práticas e nas suas avaliações tácitas ou informais. Nestas abordagens, que dão uma particular relevância aos processos e às atividades desenvolvidas, utilizam-se frequentemente dados de natureza qualitativa, promove-se a participação mais ativa dos participantes e intervenientes e assume-se claramente a subjetividade, sem receios de que esta possa retirar rigor às avaliações e aos seus resultados. As perspetivas de avaliação orientadas por uma forte *agenda social*, isto é, claramente comprometidas com processos de melhoria da vida social, económica e política das sociedades (e.g., House & Howe, 2003; Patton, 2003; Guba e Lincoln, 1989; Stake, 2006) têm subjacentes as características essenciais das abordagens que dão particular relevância aos significados que as pessoas atribuem às suas experiências e às suas práticas. A lógica, no âmbito destas abordagens e perspetivas, está mais diretamente relacionada com a argumentação, com a procura da credibilidade e da plausibilidade, assumindo que a verdade, a certeza, as conclusões definitivas, não são alcançáveis numa avaliação, tal como acontece, em muito boa medida, nas ciências físicas ou na matemática (ver, por exemplo, House, 2000; House e Howe, 1999; Howe, 2003). Mas nunca será demais sublinhar que, ainda assim, as avaliações contribuem para que se possa conhecer e compreender bem uma diversidade de objetos e de processos com eles relacionados (e.g., aprendizagens, desempenhos profissionais, qualidade da formação, projetos e programas educacionais) e, dessa forma, podem contribuir para os regular, reorientar e melhorar.

Relativamente a estas duas classes de abordagens – *A Avaliação Baseada em Critérios* e *a Avaliação Baseada nas Práticas e/ou na Experiência Pessoal* – seria eventualmente limitador considerá-las inconciliáveis e dicotômicas. É certo que os fundamentos de uma e de outra pressupõem diferentes epistemologias, diferentes lógicas de produção de conhecimento, mas também é certo que muito dificilmente poderemos produzir boas e significativas descrições da realidade se ignorarmos uma delas, seja ela qual for. Ambas são necessárias e ambas devem ser utilizadas inteligente e articuladamente. A este propósito diz-nos Robert Stake, que aqui traduzo livremente: “A experiência não garante uma interpretação correta mas a sua ausência é praticamente uma garantia para uma interpretação incorreta” (Stake, 2006, p. 134). Sendo a avaliação, no essencial, um processo deliberado e sistemático de recolha de informação que permite a formulação de juízos acerca da qualidade de um dado objeto, parece óbvio que é relevante produzir descrições tão nítidas quanto possível desse mesmo objeto. E, por isso mesmo, interessa que se articulem perspetivas e visões que aquelas abordagens proporcionam sobre as realidades que se pretendem conhecer e compreender. Olhar para a realidade sob diferentes pontos de vista pode proporcionar leituras mais profundas e mais compreensivas dessa mesma realidade.

O que parece ser importante sublinhar nesta altura é a existência de um domínio do conhecimento – a Avaliação - que informa as ações dos seres humanos em matérias como a avaliação externa das aprendizagens dos alunos. Porém, a avaliação não é, por natureza, uma ciência exata e, nesse sentido, não produz, normalmente, resultados exatos ou certos. No entanto, espera-se que esses resultados sejam reconhecidamente úteis, plausíveis e credíveis. E que sejam eticamente adequados e tão rigorosos quanto possível. A avaliação, como construção social que é, implica negociação e interação com uma diversidade de intervenientes que estão mais ou menos interessados nos seus processos e resultados e que podem ser mais ou menos afetados por eles. Consequentemente, a subjetividade faz parte de qualquer processo de avaliação e é necessário compreender que isso não significa que seja pouco rigorosa ou pouco credível e, muito menos, que seja totalmente arbitrária. A avaliação tem método, tal como a investigação, e, consequentemente, utiliza procedimentos e técnicas próprios desse método que lhe conferem rigor e que permitem credibilizar os seus resultados.

Pense-se, por exemplo, no processo de definição dos critérios que orientam a elaboração de um exame. Na verdade, é um processo que culmina em decisões que

são socialmente construídas pois decorrem de discussões entre uma diversidade de intervenientes que, naturalmente, terão perspetivas sociais, políticas e pedagógicas diferentes. É, de facto, um processo eminentemente subjetivo ou, se quisermos, intersubjetivo, e não será por isso que o exame será menos rigoroso. Muito pelo contrário. Deste modo, nunca será demais referir que a avaliação não é uma mera medida pois inclui práticas sociais que são muito exigentes e sofisticadas, incluindo componentes subjetivas que não se podem ignorar. Mas é igualmente importante sublinhar que, quando estamos a falar da avaliação no âmbito de sistemas sociais nos domínios da saúde, da educação, da segurança social e da justiça, temos que pensar que o sucesso de qualquer sistema credível de avaliação depende, antes do mais, do seu sucesso político. Um sistema de avaliação das aprendizagens num dado país será, antes do mais, uma questão política que, naturalmente, terá relações mais ou menos complexas, mais ou menos evidentes, com o conhecimento, com as teorias. As questões técnicas e tecnológicas, que são de grande relevância na produção das avaliações externas (ver, por exemplo, Madaus, Russell e Higgins, 2009), são obviamente equacionadas na sequência das decisões políticas.

Acerca Da Inserção Da Avaliação Das Aprendizagens No Sistema Educativo Português

A avaliação do que os alunos sabem e são capazes de fazer, vulgo avaliação das aprendizagens, é uma área prática, ou, como alguns autores preferem, uma área aplicada do domínio do conhecimento da avaliação. Esta é uma constatação evidente mas da qual é preciso retirar as devidas ilações. De facto, é importante compreender que a avaliação das aprendizagens está indelevelmente associada à avaliação como domínio do conhecimento e, conseqüentemente, à discussão acerca das suas questões críticas fundamentais (e.g., fundamentos, propósitos, funções, adequação ética). Isto significa que as decisões que se tomam relativamente à chamada avaliação dos alunos ao nível do sistema educativo, ao nível da escola e ao nível da sala de aula não podem deixar de estar vinculadas a perspetivas e abordagens teóricas que se têm vindo a construir no domínio da avaliação em geral. Dito de outra forma, as opções que se tomam no domínio da avaliação interna ou externa das aprendizagens, a qualquer nível, estão sempre vinculadas a enquadramentos teóricos e conceptuais que, invariavelmente, incluem uma das seguintes possibilidades: a) uma ou mais abordagens baseadas em critérios; b) uma ou mais abordagens baseadas nas práticas

e/ou na experiência das pessoas; e c) uma qualquer combinação entre as abordagens referidas em a) e em b).

A reflexão acerca dos modelos de utilização das avaliações internas e externas nas políticas públicas de educação não pode deixar de ter presente a discussão anterior mas, como é evidente, ela tem essencialmente a ver com o que pretendemos que as escolas sejam agora e no futuro. É, por isso, um debate que está sempre imbuído de valores, perspetivas, ideologias e políticas acerca da educação, das escolas e, em última análise, acerca da sociedade em que se pretende viver.

Em Portugal, por exemplo, houve uma significativa evolução nos últimos 50 anos no que se refere ao papel que a avaliação das aprendizagens dos alunos deveria ter no sistema de ensino. No estertor do Estado Novo, a avaliação interna era essencialmente baseada nos chamados *pontos* (testes) e nas *chamadas orais* e o seu principal objetivo era classificar os alunos para decidir acerca da sua progressão académica. A avaliação externa, consubstanciava-se em exames nacionais administrados nos finais dos diferentes ciclos de escolaridade a praticamente todas as disciplinas do currículo. Os exames tinham uma ponderação de 100% e, assim, a avaliação interna, da responsabilidade dos professores e das escolas, não era objeto de qualquer valorização para efeitos de certificação, servindo exclusivamente para determinar quais os alunos que transitavam de ano e, nos anos terminais, quais os que poderiam ir a exame. A certificação era exclusivamente determinada pela avaliação externa. Estávamos então perante uma avaliação consistente com um sistema de ensino essencialmente discriminatório e fortemente seletivo ao serviço dos propósitos do regime ditatorial sob o qual então se vivia em Portugal (Fernandes, 2014).

A partir de 1974 iniciou-se inevitavelmente a construção de um sistema de avaliação das aprendizagens dos alunos consistente com os valores de uma sociedade democrática e com um sistema educativo mais inclusivo. A publicação do Despacho 98-A/92, de 19 de junho, cerca de 18 anos após a instauração da democracia, “marcou indelevelmente a história da avaliação das, e para as, aprendizagens, pois consagrou princípios, conteúdos e métodos consentâneos com a ideia de que, antes do mais, a avaliação interna deve estar orientada para melhorar o ensino dos professores e as aprendizagens dos alunos.” (Fernandes, 2014, p. 255). O Despacho, referente ao ensino básico, enunciou três princípios que ainda hoje estruturam a arquitetura do

sistema de avaliação dos alunos no nosso país: a) o predomínio da avaliação formativa nas avaliações internas, tendo em vista a regulação e a melhoria do ensino e das aprendizagens; b) o predomínio da avaliação interna, assumindo os professores e as escolas um relevante papel nos processos de avaliação dos alunos; e c) o predomínio da lógica de ciclo de escolaridade, e não de ano, no que se refere às tomadas de decisão relativas à progressão académica dos alunos. Pode afirmar-se que, no plano do currículo proposto, o sistema de avaliação dos alunos do ensino básico obedece, ainda hoje, a estes princípios, apesar da introdução progressiva de exames de Matemática e de Língua Portuguesa nos anos terminais dos ciclos da escolaridade básica. Pode afirmar-se ainda que, as mudanças de governos e de políticas e dos correspondentes discursos e práticas não têm alterado, em termos legislativos e normativos, um sistema que se pode considerar progressivo e cujos princípios fundadores são, nas grandes linhas, consistentes com as recomendações constantes na literatura nacional e internacional.

Apesar do sistema de avaliação constante no currículo proposto se basear em princípios pedagógicos, em teorias do currículo, das aprendizagens e da avaliação que merecem um alargado consenso na comunidade educativa, a verdade é que persistem os problemas endémicos do sistema educativo português. Principalmente a utilização quase exclusiva da avaliação para classificar os alunos, os níveis anormalmente elevados de retenção e os resultados ainda modestos em provas de avaliação externa (nacionais ou internacionais), sobretudo quando as questões exigem a mobilização, a integração e a aplicação de conhecimentos. Avaliar para classificar, para selecionar ou para certificar continuam a ser as preocupações dominantes. Por isso, as políticas públicas de educação deveriam estar mais focadas na real introdução de práticas sistemáticas de avaliação formativa nas salas de aula para apoiarem os alunos a aprenderem melhor, com mais profundidade e significado (Fernandes, 2005, 2006a, 2007, 2009).

Como já se referiu, o impacto do Despacho 98-A/92 foi muito significativo na sociedade portuguesa e, muito particularmente, nos meios escolares e académicos, porque consagrou de uma forma nunca antes vista um conjunto de princípios que tinha implicações profundas nas formas de encarar o ensino, a aprendizagem e a avaliação. Dir-se-ia que tais princípios obrigavam a repensar a própria escola e os modos de trabalhar nas salas de aula. Por isso, a avaliação das aprendizagens assumiu

uma centralidade até aí nunca vivida no sistema educativo português, tendo sido objeto de um conjunto de medidas de política que, no essencial, consistiram na formação de professores, em sessões de esclarecimento e na produção e divulgação de materiais de apoio. A maioria daquelas medidas foram concebidas e desenvolvidas pelo Instituto de Inovação Educacional (IIE) que, mais tarde, viria a ser extinto. Gerou-se então uma dinâmica, pouco usual nos meios da educação, em torno da questão da avaliação das aprendizagens que se pode enquadrar no ímpeto reformista do ministro Roberto Carneiro, na sequência da publicação da Lei de Bases do Sistema Educativo (Lei n.º 46/86 de 14 de outubro). Efetivamente a Lei de Bases definiu uma *Agenda* para o desenvolvimento da educação em Portugal e Roberto Carneiro, na vigência do XI Governo Constitucional, tomou um conjunto de medidas que abriu caminho à modernização e desenvolvimento do sistema democrático de ensino, como é o caso das que foram tomadas no domínio da avaliação das aprendizagens dos alunos. Alguns aspetos dessa dinâmica, eventualmente por efeito de inércia, ainda perduraram no sistema por algum tempo, já na vigência do XII Governo Constitucional, sobretudo com a equipa do ministro Couto dos Santos.

O impacto do Despacho 98-A/92 foi de tal modo significativo na sociedade portuguesa que acabou por deixar na penumbra o Despacho n.º 338/93, de 21 de outubro. Este Despacho consagrou essencialmente os mesmos princípios mas o seu verdadeiro alcance e, se quisermos, a sua novidade, foi a reintrodução dos exames nacionais no ensino secundário com o duplo propósito de certificação e de acesso ao ensino superior. Assim, a partir de 1995/1996, o sistema de avaliação das aprendizagens dos alunos, através dos exames nacionais, não mais deixou de marcar profundamente o sistema educativo português ainda que o seu peso para efeitos de certificação fosse apenas de 30%, em contraste com o peso de 70% atribuído à avaliação interna, da responsabilidade integral dos professores e das escolas. Esta relação foi mantida na introdução progressiva dos exames de Matemática e de Língua Portuguesa, primeiro no 9.º ano de escolaridade, em 2004/2005, e, mais recentemente, no 4.º e no 6.º anos de escolaridade, 2012/2013 e 2011/2012, respetivamente. Para efeitos de acesso ao ensino superior aquelas ponderações são de 50% para a avaliação interna e 50% para a avaliação externa nas disciplinas específicas.

A introdução dos exames acabou por induzir, pela primeira vez em Portugal, a criação de uma estrutura com os propósitos de conceber, produzir e desenvolver instrumentos

de avaliação externa das aprendizagens (e.g., provas aferidas, exames). Efetivamente, através do Decreto-Lei n.º 229/97 de 30 de agosto, foi criado o Gabinete de Avaliação Educacional (GAVE) que viria a ser extinto e substituído pelo Instituto de Avaliação Educativa, I.P. (IAVE, IP), criado pelo Decreto-Lei n.º 102/2013 de 25 de julho. No essencial o IAVE cumpre as mesmas funções do extinto GAVE. Estas estruturas vieram profissionalizar a conceção e a elaboração dos instrumentos de avaliação externa e permitiram melhorar significativamente a sua qualidade. Porém, isto não significa que não subsistam problemas para resolver (e.g., questões relacionadas com a validade e fiabilidade das provas, com a fiabilidade entre corretores, com a análise estatística dos resultados).

As provas aferidas previstas nos Despachos que se vêm referindo acabaram por ser aplicadas regularmente no ensino básico a partir do ano 2000, sempre nas disciplinas de Matemática e de Língua Portuguesa. O potencial destas provas para que o país pudesse ter acesso a dados de qualidade sobre as aprendizagens dos alunos num espetro alargado de disciplinas era real e significativo mas foi completamente desperdiçado. Tudo isso porque as provas, a sua natureza, os seus propósitos e a sua inserção no sistema educativo e no sistema de avaliação das aprendizagens, nunca chegaram a ser realmente compreendidos por uma diversidade de intervenientes com responsabilidades no processo. Politicamente falando, nunca chegou a haver verdadeiramente uma ideia para as provas aferidas que as integrasse num todo coerente de contribuições para a avaliação dos alunos, das escolas e do sistema. Nunca se criou um programa de investigação para descrever, analisar e interpretar os dados obtidos e que fosse capaz de intervir no sentido de fazer recomendações que apoiassem as políticas públicas de educação, as escolas e os professores. Os poucos relatórios que se elaboraram não tiveram nunca esse propósito fundamental. Nestas condições, as provas foram sendo desvalorizadas acabando por ser gradualmente substituídas pelos atuais exames. Não cabendo no âmbito deste artigo elaborar sobre esta matéria, pode o leitor consultar uma detalhada discussão sobre as provas aferidas em Fernandes (2005).

Portugal aderiu aos estudos internacionais de avaliação das aprendizagens em 1991, através da participação no *International Assessment of Educational Progress II* (IAEP II) promovido pelo *Educational Testing Service* (ETS) dos Estados Unidos da América. Tratou-se de um estudo no domínio da Matemática e das Ciências, em que

participaram 20 países que testaram conhecimentos de alunos com 13 anos de idade que frequentavam a escola. Na segunda metade dos anos 90, Portugal participou no *Third International Mathematics and Science Study* (TIMSS), da responsabilidade da *International Association for the Evaluation of Educational Achievement* (IEA), destinado a alunos de 9 anos (3.º ou 4.º anos de escolaridade) e de 13 anos (7.º ou 8.º anos de escolaridade) que foram testados nos domínios curriculares da Matemática e das Ciências. Além disso, em 2000, o nosso país também participou na primeira edição do *Programme for International Student Assessment* (PISA), da responsabilidade da OCDE. De então para cá a participação portuguesa nos estudos internacionais promovidos pela IEA e pela OCDE tem sido uma constante. Mas, como já referi em trabalhos anteriores (e.g., Fernandes, 2005, 2007, 2009), durante um período relativamente longo (1991-2003), esta participação foi algo inconsequente e parece nunca ter obedecido a qualquer política deliberada e com propósitos claros. Uma das consequências da ausência de uma política de participação nas avaliações internacionais foi a forma como os resultados foram sendo apresentados pela comunicação social e apreendidos pela sociedade: uma espécie de classificação e ordenação de países como se de um campeonato de futebol se tratasse. Outra consequência foi a ausência de qualquer análise mais fina e contextualizada dos dados referentes a Portugal que permitisse compreender melhor a distribuição dos resultados. Ainda outra teve a ver com o facto de nunca ter sido claro para largos setores da sociedade e provavelmente para uma diversidade de responsáveis políticos, que a amostragem, sobretudo nas primeiras séries do PISA, prejudicou sempre a participação portuguesa. Repare-se que as amostras chegaram a incluir alunos com 15 anos que frequentavam o 5.º ano de escolaridade quando a grande maioria dos jovens da mesma idade dos outros países europeus participantes frequentavam o 10.º ou o 11.º anos de escolaridade. Finalmente, apesar da clara evolução verificada na construção de itens e de testes e na sua administração, na elaboração e utilização de critérios de correção, na análise de dados e na contextualização das avaliações, persistem problemas que ainda não estão resolvidos e que têm que ser tidos em conta. Beaton, Postlethwaite, Ross, Spearritt & Wolf (2000), Goldstein (1996, 2004), Kellaghan (2003), Kellaghan & Grisay (1995) e Riley & Torrance (2003) são autores de referência que têm sinalizado problemas que persistem nos estudos internacionais de avaliação e que aconselham uma leitura cuidada e tão informada quanto possível dos seus resultados.

Em suma, durante aquele período (1991-2003), não se conhecem reais e significativas consequências políticas, pedagógicas ou outras da participação portuguesa nos estudos internacionais de avaliação das aprendizagens. Para além de uma certa transferência de tecnologia relacionada com os processos, técnicas e procedimentos inerentes à conceção, administração e tratamento de dados, não se conseguem vislumbrar quaisquer relações que se tivessem estabelecido e assumido politicamente. Por isso, é razoável afirmar-se que não houve propriamente uma visão política para a avaliação das aprendizagens dos alunos e, talvez mais importante, uma política global e integrada que a enquadrasse na melhoria do desempenho do sistema educativo português, onde tivemos e continuamos a ter problemas sérios, apesar das melhorias já alcançadas. Porém, como se verá na discussão que se segue, esta situação passou a ser encarada de forma distinta nos anos que se seguiram.

Acerca Das Relações Entre As Políticas de Avaliação Externa E A Melhoria Das Aprendizagens Dos Alunos

A partir do segundo lustro do século XXI as políticas públicas para a educação em Portugal assumiram, de uma forma até aí nunca vista, a participação nas avaliações internacionais com claro destaque para o PISA. Parece ter-se partido do princípio de que era possível melhorar os resultados dos alunos nessas avaliações através de um conjunto de medidas de política destinadas a melhorar a qualidade do ensino e das aprendizagens. Efetivamente, ao longo de vários anos, foi concebido e posto em prática um conjunto de *programas para a qualidade* (Rodrigues, 2010) abrangendo uma grande variedade de domínios (e.g., Plano de Ação para a Matemática, Plano Nacional de Leitura, Escola a Tempo Inteiro, Avaliação Externa das Escolas). Todas estas medidas foram propositada e deliberadamente postas em prática para melhorar o desempenho do sistema educativo através do envolvimento e da participação informada dos principais intervenientes e através de uma diversidade de ações (e.g., formação dos professores, equipamento das escolas). Parece ser legítimo considerar que houve uma relação mais ou menos estreita entre as políticas e as medidas delas decorrentes e a avaliação externa das aprendizagens preconizada pela OCDE através do PISA.

Mas também é importante sublinhar que, das medidas que então foram postas em prática, algumas delas (e.g., no âmbito da Matemática, das Ciências e da Língua

Portuguesa) foram reconfigurações, aprofundamentos e melhoramentos de medidas que já tinham sido postas em prática anteriormente, nomeadamente no âmbito da chamada revisão curricular do ensino secundário ocorrida no final dos anos 90 (Fernandes, 2006b). Neste tempo, como acima já se referiu, não havia qualquer política pensada e deliberada relativamente à participação de Portugal no PISA. O poder político ou, pelo menos, uma parte dele, parecia dar mostras de alguma relutância em relação a tal participação e, no entanto, aprovou medidas de ação para melhorar o ensino e as aprendizagens dos alunos em diferentes domínios disciplinares do currículo. O que se pretende aqui dizer com esta observação é que o país tem desenvolvido, em diferentes tempos e contextos, políticas educativas que não são necessariamente decorrentes da sua participação nos estudos internacionais ou das pressões de organismos internacionais.

Voltando às medidas de política empreendidas a partir de 2005, o chamado *Plano de Ação para a Matemática* é um exemplo de como foi possível articular os esforços dos investigadores, dos formadores, dos professores, das escolas e da administração, no sentido de alterar positiva e significativamente as formas de ensinar e de aprender como ilustram as avaliações que então se realizaram (e.g., Fernandes, Borralho, Vale, Gaspar e Dias, 2011; Fernandes, Vale, Borralho, e Cruz, 2010). Efetivamente, tal como é referido nestes relatórios de avaliação, geraram-se dinâmicas de trabalho nas escolas, na participação dos professores e na cooperação entre instituições do ensino básico e do ensino superior que favoreceram a criação de ambientes mais propícios ao desenvolvimento da qualidade do ensino e das aprendizagens.

O que, de algum modo, pode ser considerado surpreendente nas políticas públicas que então foram postas em prática é o facto da participação numa avaliação externa internacional ter sido um motivo deliberadamente assumido para melhorar o ensino e a aprendizagem e o desempenho do sistema educativo. Ainda que não seja legítimo estabelecer relações de causa e efeito entre as medidas postas em prática e os resultados dos alunos nas edições de 2009 e 2012 do PISA, a verdade é que tais resultados melhoraram significativamente. De igual modo, ao contrário das expectativas menos positivas que se poderiam ter em relação aos resultados dos alunos nos exames de Matemática e de Língua Portuguesa no final do ensino básico, a verdade é que, entre 2008 e 2011, a maioria dos alunos teve notas positivas nesses mesmos exames (PORDATA, 2014).

Houve algo de novo que merece ser analisado e aprofundado pois, tanto quanto é possível compreender através dos estudos de avaliação então desenvolvidos (e.g., Fernandes *et al.*, 2010, 2011) os professores não foram propriamente mobilizados para preparar os alunos para o PISA ou para os exames, nem as escolas foram transformadas numa espécie de *centros de explicações*. Parecia haver uma ideia acerca do que a escola pública podia ser, acerca dos meios existentes e acerca das medidas que era necessário pôr em prática. Como referi mais acima é necessário ter uma visão integrada e global dos problemas educativos e, acima de tudo, tal como nos é dito por Barnett (2009), ter pensamento relativamente ao *projeto de inteligência* que o currículo deve ser e também relativamente à escola que se pretende ter e que se quer construir. Só a partir dessa elaboração de ideias se poderão pôr em prática medidas de política que permitam combater alguns dos principais e endémicos males do sistema educativo português tais como a retenção e o abandono escolar. Por isso mesmo, em função do que se pôde constatar naquele período de tempo, pode fazer sentido pensar-se numa nova geração de políticas públicas para a educação, baseada em medidas concretas, focadas nas aprendizagens dos alunos e na cooperação e colaboração privilegiadas com uma diversidade de intervenientes. Deste modo, parece ser possível lidar mais eficazmente com a complexidade dos problemas.

Apesar do que acabou de ser discutido, continua a fazer sentido e a ser necessário refletir acerca de questões tais como: Em que medida é que as avaliações externas, nacionais ou internacionais, asseguram um ensino e aprendizagens de elevada qualidade? Em que medida é que tais avaliações contribuem para resolver uma variedade de problemas que afetam os sistemas educativos (e.g., qualidade do ensino, qualidade das aprendizagens, oportunidades para aprender, clareza acerca do que é importante ensinar e aprender)? Embora se tenha apresentado um exemplo sugerindo que a conjugação das avaliações externas com um conjunto global e integrado de medidas orientadas para melhorar as aprendizagens e o ensino teve efeitos positivos no desempenho do sistema educativo, as respostas àquelas questões estão longe de estar bem estabelecidas.

As avaliações externas ou se quisermos, os exames, têm propósitos que não estão propriamente focados na melhoria das aprendizagens tais como: a) controlar, tendo em vista garantir que os conteúdos previstos no currículo são ensinados e supostamente aprendidos por todos os alunos; b) monitorizar, relacionado com a

responsabilização e a chamada prestação de contas através dos resultados obtidos pelos alunos; c) certificar; e d) selecionar, como é, entre nós, o caso no acesso ao ensino superior por parte dos jovens que concluem o ensino secundário (e.g., Kellaghan e Madaus, 2003). Obviamente que não podemos ignorar algumas das vantagens dos exames, partindo do princípio de que estão bem feitos, tais como: a) o efeito moderador que têm sobre as avaliações internas; b) a indução de práticas inovadoras de ensino e de avaliação; c) a contribuição para se compreender melhor o desempenho dos sistemas educativos, melhorando a tomada de decisões; d) a informação que podem proporcionar às escolas, aos professores e aos alunos relativamente ao que é importante ensinar e aprender; e e) a mobilização das escolas e dos professores para a eventual necessidade de reverem os seus processos de trabalho, nomeadamente no que se refere aos projetos educativos e à organização e desenvolvimento do ensino.

Porém, como de algum modo já atrás se referiu, as avaliações externas das aprendizagens estão igualmente associadas a um conjunto de desvantagens das quais se destacam aqui o chamado *estreitamento* ou *afunilamento* do currículo, que consiste na tendência para que o ensino se centre no que supostamente sai nos exames. Ou seja, só é valorizado e, por isso, só é ensinado, o que é examinado. Consequentemente, as disciplinas que não são examinadas perdem o seu estatuto, a sua importância e esvaziam-se. O currículo é inevitavelmente empobrecido pois domínios relevantes para a formação das crianças e dos jovens podem não ser sequer abordados ao longo da escolaridade. Outra desvantagem tem a ver com o facto de as escolas serem induzidas a prestar mais atenção aos alunos que pensam poder vir a ter sucesso nos exames em detrimento dos que têm dificuldades ou em relação aos quais as expectativas são baixas. Consequentemente, os exames podem induzir práticas discriminatórias que levam à exclusão de alunos e, em particular, dos que são oriundos de meios mais fragilizados económica e socialmente.

A este propósito, Madaus, Russell e Higgins (2009) referem que as avaliações externas são um sistema paradoxal porque, por um lado, produzem efeitos que podem ser considerados positivos e que são deliberadamente desejados (e.g., cumprimento dos programas, utilização de estratégias de ensino e avaliação inovadoras, maior empenhamento dos alunos, ensino mais orientado para questões consideradas fundamentais) mas, por outro lado, produzem igualmente efeitos colaterais, não

desejados e não intencionais com consequências francamente negativas e que podem prejudicar os alunos e as famílias (e.g., estreitamento do currículo, ensinar apenas para o exame, estudar apenas os assuntos que supostamente saem no exame, abandono escolar precoce, retenção).

Todos os defensores das avaliações externas concordam que aquilo que se ensina e aquilo que se aprende é, em muito boa medida, por elas determinado. Nesse sentido, de acordo com vários autores (e.g., Fernandes, 2005; Kornhaber, 2004; Madaus, Russell e Higgins, 2009), as escolas e os seus professores utilizam uma diversidade de estratégias para lidar com a questão dos exames tais como: a) desenvolver planos de preparação específica; b) planejar aulas adicionais; c) treinar especificamente as respostas a certas questões; d) treinar os alunos a descartarem certas opções nas perguntas de escolha múltipla; e e) exercer pressão junto dos alunos com mais dificuldades para não irem a exame ou para saírem da escola. Uma reflexão sobre estas estratégias pode mostrar que há efeitos colaterais indesejáveis e até condenáveis que as avaliações externas podem gerar. E também pode contribuir para pensar acerca da natureza e da profundidade das aprendizagens que são desenvolvidas nestas condições. Poderá eventualmente perguntar-se: em que medida é que responder acertadamente a um conjunto de perguntas constantes num exame, traduz realmente aprendizagens seguras e profundas acerca de um dado domínio do currículo?

Em suma, por um lado, tem-se verificado que os proponentes e defensores das avaliações externas como forma de medir o que os alunos supostamente aprendem, argumentam que os exames estabelecem metas claras e, assim, as escolas e os professores orientam os seus esforços para as capacidades e conhecimentos normalmente definidos no currículo. Mas, por outro lado, os críticos desta abordagem referem que o ensino orientado para a medida, baseado em exames, conduz inevitavelmente à distorção e estreitamento do currículo, empobrecendo-o, fragmentando-o e, conseqüentemente, prejudicando a aprendizagem dos alunos (e.g., Gipps, 1994; Kellaghan, 2003, Kellaghan e Madaus, 2003, Kellaghan e Greaney, 2001). Porém, nenhum destes grupos tem conseguido demonstrar cabal e irrefutavelmente as suas afirmações e, nesta situação, a discussão é mais baseada nos valores, crenças e ideologias dos diferentes intervenientes do que em qualquer racionalidade devidamente fundamentada. Assim, talvez estejamos perante um domínio em que a persuasão, mais ou menos fundamentada e baseada nas ideologias,

crenças e valores dos protagonistas, parece ter uma significativa relevância. É sabido, por exemplo, que, tradicionalmente, para largos sectores da opinião pública, as avaliações externas, nacionais e internacionais, constituem símbolos e medidas irrefutáveis da qualidade dos sistemas educativos e, concomitantemente, da qualidade dos seus professores, das suas escolas e dos seus currículos. Porém, não há, como se disse, qualquer evidência científica que comprove essa ideia, essa crença. Para uma discussão aprofundada sobre esta questão crítica vejam-se, por exemplo, para além da literatura acima referenciada, Harlen (2007), Kellaghan e Grisay (1995), Kornhaber (2004) e Stobart (2008).

Nestas condições e sabendo-se igualmente que não há instrumentos, como os exames, que sejam perfeitos, a questão põe-se em termos da ponderação que se pode fazer entre os efeitos benéficos e os efeitos indesejáveis e até prejudiciais que estão indelevelmente associados a qualquer sistema de avaliação externa. Consequentemente, o exercício que será sempre necessário fazer consiste em determinar rigorosamente, as vantagens e benefícios e as desvantagens e prejuízos do sistema que eventualmente se quer pôr em prática, para poder ser possível decidir de forma tão equilibrada e tão fundamentada quanto possível.

Como se não bastassem já os pontos críticos enunciados importa igualmente ter em consideração que as avaliações externas podem ser utilizadas com diversas vantagens para auditar os sistemas educativos mas não são normalmente consideradas adequadas para ajudarem a ensinar e a aprender melhor. Por seu lado, as avaliações internas, da responsabilidade direta dos professores e das escolas, podem, comprovadamente, melhorar significativamente o ensino e a aprendizagem mas, por razões já acima apresentadas, dificilmente poderão ser utilizadas para auditar um sistema educativo. Consequentemente, parece que ambas deverão ser inteligentemente articuladas porque, por natureza e tendo em conta o estado atual da discussão sobre estas questões, as avaliações internas e externas poderão ser complementares. Na verdade, uma análise das vantagens e desvantagens de cada um dos sistemas mostra que a adesão exclusiva e rígida a um deles terá implicações negativas a vários níveis. Por isso, apostar na sua complementaridade e na sua articulação parece poder ser o caminho mais adequado tal como, aliás, vem sendo defendido por vários autores (e.g., Kornhaber, 2004; Madaus, Russell e Higgins, 2009).

Conclusões e Reflexões Finais

Quando pensei num título para este trabalho decidi incluir a expressão *Questões Críticas De Uma Relação (Im)Possível* porque, naturalmente, pretendia apresentar e discutir alguns dilemas que surgem quando se pensa na relação entre as avaliações externas e a melhoria das aprendizagens dos alunos. Através da expressão *(Im)Possível*, pretendi desafiar o leitor quanto à possibilidade ou impossibilidade dessa relação. Assim, chegados a este ponto, espero que tenha ficado claro que tal relação será impossível se houver perspectivas que considerem que as avaliações internas ou as avaliações externas, isoladamente consideradas, são a panaceia para a resolução dos problemas de aprendizagem existentes no sistema. Se, por outro lado, reconhecermos que, cada uma delas, por si só, é insuficiente para enfrentar os problemas, porque ambas têm vantagens e desvantagens que, sob muitos pontos de vista, se complementam, então parece razoável considerar-se que tal relação pode ser possível.

Do ponto de vista do conhecimento produzido neste domínio, da teoria, se quisermos, a relação entre as avaliações externas e a melhoria das aprendizagens é possível mas é necessário considerar e pôr em prática um conjunto de ações. Limitar-me-ei, neste contexto, a referir as que me parecem mais relevantes.

A triangulação de informação é um princípio metodológico sempre aconselhável quando se pretende melhorar a qualidade da informação obtida. Neste sentido, face às limitações das avaliações internas e externas, será sempre positivo articular inteligentemente os dados obtidos através de cada uma delas. Por outro lado, de acordo com os resultados da investigação neste domínio, nenhuma decisão relativamente a questões tais como a retenção dos estudantes, a sua inserção em programas de recuperação ou de remediação e à certificação deveria ser tomada com base no resultado de um único teste ou exame externo (Kornhaber, 2004). Sabe-se que, do ponto de vista da psicometria, as decisões tomadas a partir dos resultados de vários testes são mais fiáveis pois o erro de medida é menor, mas também se sabe que uma medida deste tipo é quase impraticável ou mesmo totalmente impraticável, por razões óbvias. Eu diria que é indesejável pois iria contribuir para o exacerbamento de algumas das desvantagens das avaliações externas. Consequentemente, a articulação entre as avaliações internas e externas parece ser um caminho a aprofundar, sobretudo

porque é efetivamente através das primeiras que os alunos podem aprender com profundidade, conforme está amplamente evidenciado na literatura (ver, por exemplo, Black e Wiliam, 1998; Gardner, 2006).

O sistema existente em Portugal, no que se refere às ponderações das avaliações externas e internas para efeitos de certificação, tem-se mantido desde os anos 90 e privilegia claramente o peso das avaliações internas. Porém, as avaliações externas dificilmente poderão ser associadas à melhoria das aprendizagens se não houver, por um lado, uma aposta política clara no desenvolvimento das práticas ao nível das avaliações internas e, por outro lado, um trabalho de aprofundamento das relações entre ambas as modalidades de avaliação. Isto significa que, em princípio, as avaliações externas só poderão contribuir para melhorar as aprendizagens dos alunos se estiverem fortemente articuladas com um sistema de avaliação interna mais credível e orientado para apoiar o ensino e as aprendizagens. Assim, torna-se igualmente incontornável apostar na credibilização das avaliações internas através de sistemas delineados pelas e nas escolas que promovam a articulação de processos e estratégias de avaliação das aprendizagens que ajudem os alunos a aprender e que contribuam para melhorar a sua validade e a sua fiabilidade. No limite, um sistema com uma avaliação interna credível dispensaria a avaliação externa. Na verdade, uma das razões que justifica a utilização das avaliações externas é o facto das avaliações internas não serem suficientemente credíveis (e.g., válidas, fiáveis) para poderem ser consideradas para efeitos de responsabilização e de prestação de contas, exigências dificilmente contornáveis nas sociedades atuais. Mas, também aqui, as avaliações internas terão um papel cada vez mais significativo à medida que a sua credibilidade se for estabelecendo.

Outra questão que é necessário equacionar neste contexto prende-se com as oportunidades de aprendizagem que os sistemas educativos proporcionam, ou não, a todos os estudantes. A investigação tem mostrado que há diferenças, que podem ser bastante sensíveis, na preparação dos alunos para os exames que têm a ver com as oportunidades que lhes foram proporcionadas para aprender. Questões tais como *Os alunos tiveram oportunidades para aprender e não aprenderam? Os alunos não tiveram oportunidade para aprender?* terão que ser sempre consideradas e ponderadas. Na verdade, uma coisa é os alunos não terem estudado/trabalhado e, conseqüentemente, não terem aprendido. Outra é o sistema educativo não lhes ter

proporcionado as necessárias oportunidades para que eles pudessem aprender. Não serão seguramente as avaliações externas que irão garantir a equidade no que se refere às oportunidades para aprender. As oportunidades para aprender exigem um forte investimento em programas vários tais como o *Plano de Ação para a Matemática* e o *Plano Nacional de Leitura* acima mencionados. É dessa forma que se alargam as oportunidades de aprendizagem e se podem relacionar as avaliações externas e internas com a melhoria das aprendizagens dos alunos. Só por ingenuidade ou por deliberada intenção política se poderá crer que as avaliações externas são a fonte privilegiada para a resolução dos problemas de aprendizagem e muitos outros (e.g., ensino, motivação dos alunos, inadequação dos currículos) existentes nos sistemas educativos. É com certeza muita atrativa a ideia de resolver os problemas com uma só medida de política tal como *criar exames a sério para melhorar o sistema* ou *promover uma avaliação rigorosa e a sério*, como tem sido afirmado por responsáveis políticos. Independentemente de não se perceber muito bem o que serão *exames a sério* ou *avaliações rigorosas e a sério*, é preciso compreender que as questões sociais complexas, como é o caso da melhoria das aprendizagens dos estudantes, não se resolvem com a utilização exclusiva de uma medida ou abordagem. Exigem a construção de políticas mais diversificadas, mais pensadas e fundamentadas que deverão incluir uma diversidade de formas de avaliação educacional.

As avaliações externas só poderão contribuir efetivamente para melhorar o ensino e a aprendizagem e o desempenho dos alunos, se estiverem plenamente integradas nos sistemas educativos e se, através delas, forem estabelecidos padrões claros a partir dos quais professores e alunos possam trabalhar. Assim, as avaliações externas terão que estar focadas no desenvolvimento das capacidades superiores de pensamento, articuladas com as finalidades e objetivos do currículo e orientadas para induzirem práticas de ensino, de avaliação e processos de estudo mais consentâneos com as recomendações decorrentes do conhecimento produzido sobre estas questões (e.g., Black e Wiliam, 1998; Gardner, 2006; Gipps, 1994; Stobart, 2008). Assim, não é suficiente termos avaliações externas acima de qualquer suspeita do ponto de vista técnico e tecnológico, é necessário garantir que elas sejam geradoras dos efeitos que se pretendem nas aprendizagens dos alunos e no ensino dos professores.

Tal como nos disse Lauren Resnick só se obtém o que se avalia e, por isso, não se obtém o que não se avalia! Consequentemente, as avaliações externas têm que ser

continuamente melhoradas e têm que estar bem focadas no que se quer, e como se quer, que os alunos aprendem (Resnick, 1987).

Finalmente, interessa ter presente que, em países com grande tradição na utilização de avaliações externas, os psicometristas e os políticos sempre estiveram plenamente convencidos de que elas, por si sós, melhorariam a educação pública. Os factos e as investigações têm mostrado que assim não é. Por isso, é preciso pensar que as avaliações externas não podem, por si sós, resolver os problemas. Têm que ser delineadas de forma mais integrada e articulada com outros processos de avaliação, internos e externos, e apoiadas em políticas ativas materializadas em programas focados nas aprendizagens dos alunos.

Referências

- Barnett, R. (2009). Knowing and becoming in the higher education curriculum. *Studies in Higher Education*, 34 (4), 429- 440.
- Beaton, A., Postlethwaite, T., Ross, K., Spearritt, D., e Wolf, R. (2000). *The benefits and limitations of international educational achievement studies*. Paris: International Institute for Educational Planning/International Academy of Education.
- Black, P. e Wiliam, D. (2006). Assessment for learning in the classroom. In J. Gardner (Ed.), *Assessment and learning*, 9-26. London: Sage.
- Black, P. e Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5, 1, 7-74.
- Bloom, B., Hastings, J. e Madaus, G. (1971). Handbook of formative and summative evaluation of student learning. New York: Mac Graw Hill
- 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*, n.º 78, vol. 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*, vol. 16, 2, 241-261.

- Fernandes, D. (2007). A avaliação das aprendizagens no sistema educativo português. *Educação e Pesquisa*, v. 33, 3, pp. 581-600.
- Fernandes, D. (2006a). Para uma teoria da avaliação formativa. *Revista Portuguesa de Educação*, 19, 2, pp. 21-50.
- Fernandes, D. (2006b). Revisitando a revisão curricular (1997-2001): Um contributo para pensar o futuro do ensino secundário. *Educação. Temas e Problemas*, 2, pp. 129- 158.
- Fernandes, D. (2005). *Avaliação das aprendizagens: Desafios às teorias, práticas e políticas*. Cacém: Texto Editores.
- Fernandes, D., Borralho, A., Vale, I., Gaspar, A. & Dias, R. (2011). Ensino, avaliação e participação dos alunos em contextos de experimentação e generalização do novo programa de matemática do ensino básico. Lisboa: Instituto de Educação da Universidade de Lisboa.
- Fernandes, D., Vale, I., Borralho, A. e Cruz, E. (2010). Uma avaliação do processo de experimentação do novo programa de matemática do ensino básico (2008/2009). Lisboa: Instituto de Educação da Universidade de Lisboa.
- Gardner, J. (Ed.) (2006). *Assessment and learning*. London: Sage.
- Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. Londres: Falmer.
- Goldstein, H. (2004). International comparisons of student attainment: Some issues arising from the PISA study. *Assessment in Education: Principles, policy & practice*, 11, 3, pp. 319-330.
- Goldstein, H. (1996). International comparisons of student achievement. In A. Little e A. Wolf (Eds.), *Assessment in transition: Learning, monitoring and selection in international perspective*, pp. 58-87. Oxford: Pergamon.
- Guba, E. e Lincoln, Y. (1989). *Fourth generation evaluation*. Newbury Park: CA: Sage.
- Harlen, W. (2007). *Assessment of learning*. London: Sage.
- House, E. (2000). *Evaluación, ética y poder* (Tercera Edición). Madrid: Morata.
- House, E. e Howe, K. (2003). Deliberative democratic evaluation. In T. Kellaghan e D. Stufflebeam (Eds.), *International handbook of educational evaluation*, 79-102. Dordrecht: Kluwer.
- House, E. e Howe, K. (1999). *Values in education and social research*. London: Sage.
- Howe, K. (2003). *Closing methodological divides: Toward democratic educational research*. Dordrecht: Kluwer.
- 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.
- Madaus, G., Russell, M. e Higgins, J. (2009). *The paradoxes of high stakes testing: how they affect students, their parents, teachers, principals, schools, and society*. Charlotte, NC: Information Age Publishing.
- Patton, M. (2003). Utilization-focused evaluation. In T. Kellaghan e D. Stufflebeam (Eds.), *International handbook of educational evaluation*, 223-244. Dordrecht: Kluwer.
- Pordata (2014). Base de dados –Portugal contemporâneo (Educação). Retirado de <http://www.pordata.pt/Tema/Portugal/Educacao-17> em 12 de outubro de 2014.
- Resnick, L. (1987). *Education and Learning to Think*. Washington, DC: The National Academies Press.
- Riley, K. & Torrance, H. (2003). Big change questions. *Journal of Educational Change*, 4, 419-425.
- Rodrigues, M. L. (2012). Os desafios de política de educação no século XXI. *Sociologia, Problemas e Práticas*, 68, pp. 171-176.
- Rodrigues, M. L. (2010). *A escola pública pode fazer a diferença*. Coimbra: Almedina.
- Scriven, M. (1967). *The methodology of evaluation* (AERA Monograph series on curriculum evaluation, No. 1). Chicago IL: Rand McNally.
- Stake, R. (2006). *Evaluación comprensiva y evaluación basada em estándares*. Barcelona: Graó.
- 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.
- Stufflebeam, D. (2000). Foundational models for 21st century program evaluation. In D. Stufflebeam, G. Madaus e T. Kellaghan (Eds.), *Evaluation models: viewpoints on educational and human services evaluation* (2nd. Edition), 33-83. Dordrecht: Kluwer.