

Avaliação Externa e Qualidade das Aprendizagens



CNE
CONSELHO
NACIONAL DE
EDUCAÇÃO

Seminários
e Colóquios

AVALIAÇÃO EXTERNA
E QUALIDADE DAS
APRENDIZAGENS



As opiniões expressas nesta publicação são da responsabilidade dos autores e não refletem necessariamente a opinião ou orientação do Conselho Nacional de Educação.

Título: Avaliação Externa e Qualidade das Aprendizagens

[Textos do Seminário realizado no CNE a 1 de abril de 2014]

Edição: Conselho Nacional de Educação

Direção: José David Justino (Presidente do Conselho Nacional de Educação)

Coordenação: Manuel Miguéns (Secretário-Geral do Conselho Nacional de Educação)

Organização e edição: Antonieta Lima Ferreira

Coleção: Seminários e Colóquios

Edição Eletrónica: dezembro de 2014

ISBN: 978-972-8360-88-7

© CNE – Conselho Nacional de Educação
Rua Florbela Espanca – 1700-195 Lisboa
Telefone: 217 935 245
Endereço eletrónico: cnedu@cnedu.pt
Sítio: www.cnedu.pt

SUMÁRIO

Nota prévia	6
<i>Secretário-Geral do Conselho Nacional de Educação</i> Manuel Miguéns	
Abertura	14
<i>Presidente do Conselho Nacional de Educação</i> José David Justino	
<i>Secretário de Estado do Ensino e da Administração Escolar</i> João Casanova de Almeida	
Conferência	
Avaliações externas e melhoria das aprendizagens dos alunos: questões críticas de uma relação (im)possível	21
Domingos Fernandes	
PAINEL	
Avaliação externa e melhoria das aprendizagens: uma relação compatível?	50
Carlinda Leite João A. Lopes Francisco Vargas Valter Lemos	
PAINEL	
Avaliação externa e práticas docentes	77
Aldina Lobo Paula Serra & Cecília Galvão	

PAINEL

O que fazer com os resultados da avaliação externa? 109

Helder Sousa
Luís Santos
Rui Santos
Júlia Tainha

ENCERRAMENTO

180

Presidente do Conselho Nacional de Educação
José David Justino

Secretário de Estado do Ensino Básico e Secundário
João Grancho

NOTA PRÉVIA

Manuel Miguéns¹

As questões relativas à avaliação têm sido, ao longo das últimas décadas, motivo de debate na sociedade portuguesa com inevitável reflexo nos trabalhos desenvolvidos pelo Conselho Nacional de Educação. O estudo, a reflexão, a audição de académicos e especialistas e o debate sobre avaliação ganham agora nova dinâmica e mereceram atenção redobrada deste Conselho.

O seminário *Avaliação Externa e Qualidade das Aprendizagens*, realizado em abril de 2014, procurou discutir e partilhar diferentes perspetivas sobre o impacto da avaliação externa (vulgo exames) na qualidade das aprendizagens dos alunos. Com a generalização das provas finais e dos exames nacionais a todos os ciclos de ensino, concretizada em 2013, decidiu o Conselho Nacional de Educação acompanhar o desenvolvimento deste processo e sobretudo ouvir os especialistas na matéria, tendo em vista a clarificação de conceitos e de finalidades, assim como a identificação das relações entre as várias modalidades de avaliação.

Na intervenção de abertura, o Presidente do CNE, reconhecendo que o nosso sistema de avaliação externa tem vindo a melhorar, afirma que os exames não são a questão fundamental. O importante é sabermos como a escola ensina e como os alunos aprendem. O centro decisivo é a qualidade do ensino e a qualidade das aprendizagens

Na conferência de abertura deste seminário, Domingos Fernandes discute a relação entre a avaliação externa dos alunos e a qualidade das aprendizagens, referindo alguma da literatura mais recente sobre a temática. Considera esta relação como impossível sempre que a perspetiva dominante considere as avaliações externas como a “panaceia” para a resolução de problemas de aprendizagem no sistema. De entre as ações que propõe contam-se a credibilização das avaliações internas, a integração plena das avaliações externas no sistema de ensino (por exemplo, com as avaliações externas a contribuírem para a definição de padrões claros) e a assunção da complementaridade entre ambas.

¹ Secretário-Geral do Conselho Nacional de Educação

Na abertura do painel sobre a compatibilidade da relação entre avaliação externa e a melhoria das aprendizagens, Carlinda Leite enfatiza as ações que promovem uma formação global e privilegiam a socialização das crianças e dos jovens para uma vivência positiva da cidadania, em vez de se acantonarem na preparação técnica dos alunos apenas para serem bem-sucedidos nos exames. No entanto, reconhece que a informação obtida através das provas nacionais permite sustentar balanços do sistema educativo e apontar caminhos às escolas, aos professores e aos alunos, com vista à melhoria da qualidade das aprendizagens.

A importância da divulgação pública dos resultados, a necessidade de não se estabelecerem relações de causalidade entre avaliação externa e melhoria das aprendizagens e de se ponderarem os custos dos programas de avaliação externa, são os pontos focais do texto que João Lopes apresenta.

Francisco Vargas, por sua vez, centra a discussão na adequação das provas externas para medir os conhecimentos e capacidades, evitando dar relevo ao “sim” ou “não” à avaliação externa. Defende, deste modo, que o importante é que as avaliações sejam capazes de considerar diferentes estilos de aprendizagem, inteligências múltiplas, capacidades superiores de pensamento, para que possam ser uma boa ferramenta.

Valter Lemos considera que “a familiaridade provoca a cegueira” e avança um claro “sim” à avaliação externa que, no seu entender, garante confiança social e responde à necessidade de regulação. E propõe que, a instrumentos diversos correspondam diferentes objetivos: exames para certificação social; provas aferidas para monitorização de programas, atores e contextos; testes internacionais para avaliação de políticas.

Aldina Lobo abre o painel sobre avaliação externa e práticas docentes apresentando um estudo sobre o modo como as práticas dos professores podem adaptar-se de forma a garantir a melhor convivência entre aprendizagens de qualidade e os resultados da avaliação externa. Em “Vestir a camisola do exame e a da prática letiva”, a autora defende que é possível fazer coabitar a avaliação formativa e a sumativa (o apoio à aprendizagem e a preparação para exame), mesmo com alunos que serão sujeitos a exames nacionais.

Para Graça Ventura a avaliação externa tem impactos positivos, como por exemplo, o aumento do trabalho colaborativo nas escolas e uma gestão responsável do currículo, mas também negativos, como sejam o estreitamento do currículo (desvalorização das disciplinas sem exame), o aprofundamento inadequado de conteúdos (que afasta alunos com mais dificuldades) ou a padronização de instrumentos e metodologias.

Paula Serra e Cecília Galvão abordam a ligação entre avaliação externa e avaliação formativa e defendem que as tarefas de avaliação incluídas nos referidos exames podem ser usadas pelos professores, em sala de aula, para desenvolver avaliação formativa, funcionando, assim, como instrumentos charneira, capazes de articular avaliação formativa e sumativa. Problematizam e discutem os fatores que poderão estar na origem dos maus resultados obtidos pelos alunos nestes exames e descrevem uma intervenção educativa implementada junto de professores do ensino secundário, que partindo de questões de exame, se centra no desenvolvimento de estratégias de avaliação formativa, focadas no desenvolvimento do raciocínio científico e da construção de explicações científicas.

A questão chave do painel seguinte pedia aos oradores que se debruçassem sobre o que fazer com os resultados da avaliação externa. Helder Sousa questiona o carácter público das provas de exame dado que esse facto, para além dos impactos negativos, tem implicações financeiras, de mobilização de recursos humanos e introduz sérias limitações à “comparabilidade” dos exames. Relativamente à leitura de resultados considera que o valioso manancial de informação disponível não tem sido aproveitado, que não se podem repetir os mesmos erros a cada ano e que o País não pode aceitar pacificamente que o topo da distribuição normal de resultados se situe próximo dos 50%. Aponta ainda um desafio: provas não públicas e desmaterializadas, pelo menos para o ensino básico.

Luís Santos inicia o seu texto com algumas questões de base: quais os dados e as métricas mais adequadas para a leitura e utilização dos resultados da avaliação externa? Devemos olhar apenas para as médias ou importa valorizar outras medidas, como por exemplo, as diferenças entre classificações internas e externas? Que tipo de informação nos dão os

resultados da avaliação externa sobre o desempenho dos alunos e das escolas? Considera que importa distinguir diferentes níveis de análise dos resultados – para decisores políticos, escolas e público – e propõe análises mais finas e desagregadas que forneçam informação relevante, por exemplo, sobre alunos do 1º ciclo sujeitos a períodos de acompanhamento extraordinário ou sobre alunos dos cursos profissionais.

A Rede ESCXEL, que trata e analisa um vasto conjunto de dados, incluindo os das classificações de exame dos alunos das escolas que a integram, foi apresentada por Rui Santos que explicou os procedimentos utilizados com vista a “dar uma nova camada de inteligibilidade aos resultados” com base nos três princípios que sustentam a própria rede – Padronizar, Resumir e Mapear.

Júlia Tainha encerrou este painel com o projeto e a história da escola que dirige, procurando demonstrar que os resultados dos exames podem, de facto, ser utilizados como um incentivo para melhorar a qualidade das aprendizagens dos alunos. Formulou ainda um desejo de maior autonomia, alicerçada em mais confiança nas escolas que têm trabalho feito e provas dadas.

Por fim, o Presidente do Conselho Nacional de Educação reconhece o importante papel da avaliação externa, dando relevo às limitações e aos desafios que se colocam no sentido de melhorar o atual modelo de exames, sem esquecer as aprendizagens e o desempenho dos alunos e das escolas.

Não tendo sido esgotada a questão, o CNE pretende continuar a dar o seu contributo para este debate, alargando a reflexão agora iniciada a outras vertentes, nomeadamente a da avaliação interna e a das avaliações internacionais.

ABERTURA

José David Justino¹

Este seminário faz parte das iniciativas que o Conselho Nacional de Educação se propõe realizar no âmbito do seu plano de atividades, procurando atingir a média de um seminário por mês. Nestes seminários pretende-se envolver um conjunto alargado de especialistas, direta ou indiretamente ligados à Educação, consolidando a ideia que tenho vindo a lançar de tornar o CNE numa espécie de *casa comum da Educação*, onde as pessoas podem vir e participar de forma ativa.

O seminário que escolhemos para iniciar este ciclo é dedicado à avaliação externa e qualidade das aprendizagens e, assim, começo por agradecer não só à assessoria, a todos os colaboradores que, na área técnica e na assessoria técnico-científica do Conselho, contribuíram para a realização deste encontro mas, especialmente, a duas pessoas: à Dr.^a Antonieta Lima Ferreira, que coordenou a organização do seminário, e ao Professor Domingos Fernandes, que nos deu uma ajuda inestimável, pelo que agradeço publicamente a sua colaboração e, acima de tudo, os contributos que entendeu, por bem, transmitir.

Porquê este tema? Como sabem, já ultrapassámos quase mais de quinze anos de existência de várias iniciativas visando constituir um sistema de provas nacionais. Dos exames de 12.^o ano às provas de aferição, da primeira vez que se realizaram provas e exames de 9.^o ano em 2005 (embora decretados em 2003) até à criação das novas provas nacionais no final de cada ciclo, tem-se vindo a cumprir o princípio anunciado há muito tempo de que no final de cada ciclo faz sentido ter um instrumento de avaliação externa.

Desde muito cedo que a avaliação externa, através de provas, quer sejam provas de aferição quer sejam provas de avaliação externa sob a forma de exames, sofreu uma contestação muito grande. Vemos o sistema a funcionar, mas não podemos esquecer aqueles que, na altura, lembraram com algum sentido, o que poderia ser a influência da realização de provas

¹ Presidente do Conselho Nacional de Educação.

no final de cada ciclo sobre a qualidade das aprendizagens. É, precisamente, com base neste argumento que tomámos a iniciativa de refletir em torno desta relação entre avaliação externa e qualidade das aprendizagens.

Aquilo que fizemos foi também, de algum modo, dizer que não há assuntos tabus. Ou seja, não é pelo facto de termos exames que deixamos de continuar a poder emitir críticas a esses mesmos exames ou eventualmente qualificar o modelo de exames tal como está a ser feito. Portanto, acho que tem plena justificação que possamos, ao fim destes anos, fazer esta reflexão. É fundamental olhar para o futuro fazendo o balanço desse mesmo passado. Isto não quer dizer que tenhamos de prescindir das nossas opiniões ou sacrificar as nossas posições. O que é importante aqui é conseguirmos encontrar uma solução de compromisso sobre o que é possível fazer e, acima de tudo, sobre o que é desejável fazer no futuro.

Temos um sistema de exames que tem vindo a melhorar. Quem acompanha o trabalho do atual IAVE sabe que tem havido melhorias ao longo dos vários anos. Desde que se começaram a fazer as primeiras provas e depois as provas de aferição que tem havido um acumular de conhecimento e de experiências que não deve ser ignorado e que tem permitido que se vão fazendo os necessários ajustes. Tal melhoria não quer dizer que se ponha de parte qualquer tentativa de começarmos a pensar sobre se este modelo é o mais adequado ou não. Uma coisa é fazer um exame no final da escolaridade obrigatória de 12 anos, outra coisa é ter um conjunto de provas no final de cada ciclo. E, portanto, nessa perspetiva, julgo que o debate nos irá permitir dar contributos de relevo que, acima de tudo, nos ajudarão a refletir sobre esta matéria.

Por outro lado, quer queiramos quer não, o fundamental não são os exames; o fundamental é sabermos como os miúdos aprendem e como é que a escola ensina. Ou seja, o exame vale o que vale, é mais um instrumento de avaliação, mas o centro decisivo é a qualidade das aprendizagens e a qualidade do ensino, e é sobre isso que nos devemos preocupar. A relação com as provas externas também não deve ser

ignorada porque, por vezes, diz-se que as pessoas só estudam para os exames; ora este é um argumento falacioso porque parte do princípio de que as pessoas sabem o que vai sair no exame. Se se soubesse, à partida, qual era o exame só se estudava aquela matéria; como não se sabe tem de se estudar a matéria toda. Por outro lado, há outras competências que eventualmente não são suscetíveis de se avaliarem através dos exames, mas também para isso existe a avaliação interna, que tem um papel importantíssimo. Os exames valem 30% e a avaliação interna vale 70% na classificação final.

Há uma polémica igualmente reconhecida por todos: a partir da altura que se elegem duas provas, duas disciplinas para se fazer os exames, a questão coloca-se em saber o que acontece com as outras disciplinas. Ou seja, estas ganham um estatuto ligeiramente inferior, podendo até sofrer com a concentração dos alunos, dos professores e das escolas nas disciplinas de exame, pois o que interessa é o resultado obtido nas provas. Não obstante, é bom relembrar o facto de que dominar bem a língua e dominar bem a matemática são saberes estruturantes que vão beneficiar todos os outros saberes.

Portanto, temos aqui muito por onde refletir, argumentos não faltam. É nesta perspetiva que entendo que o debate aberto e franco é sempre positivo e é precisamente para isso que vos convido, contando com a experiência de muitos anos dos muitos aqui presentes e que reconhecidamente têm dado muito do seu trabalho para que este sistema de exames possa, no fundo, frutificar, ao mesmo tempo que, em conjunto, se identificam pontos de insuficiência ou pontos de maior dificuldade para que, junto do Ministério, possamos fazer chegar a nossa mensagem. É esse o nosso papel, refletir e, com base nessa reflexão, elaborar recomendações que possam ajudar, apoiar o trabalho de quem tem a responsabilidade das políticas públicas de educação.

Portanto, Senhor Secretário de Estado, quero agradecer-lhe a sua presença, a forma como demonstrou interesse nesta iniciativa, a colaboração que tem existido sempre entre o Ministério de Educação e Ciência e o Conselho Nacional de Educação e também dizer-lhe, obviamente, que

quer consigo, quer com o Sr. Secretário de Estado do Ensino Básico e Secundário, João Grancho, logo ao final do dia, estamos sempre disponíveis e abertos para colaborar, desde que essa colaboração se traduza em ganhos substanciais e num valor acrescentado para a qualidade da Educação em Portugal. Muito obrigado pela vossa atenção.

João Casanova de Almeida¹

Senhor Presidente do Conselho Nacional de Educação, meu caro amigo Professor David Justino, saúdo todos aqueles que hoje vão contribuir, com as suas comunicações, para a reflexão que vai ter lugar, todos aqueles que, com a sua participação no debate, vão enriquecer essas reflexões e vão permitir perceber a importância que tem a organização de seminários como este, no Conselho Nacional de Educação, para, como dizia o Professor David Justino, podermos ter todos à nossa disposição reflexões que permitam perceber da maior ou menor importância, do grau de importância, de cada um dos fatores que contribuem para a qualidade das aprendizagens dos nossos alunos. Por isso, a primeira palavra de agradecimento é para o Conselho Nacional de Educação, na pessoa do seu presidente, pelas iniciativas que tem tomado e pelo anúncio que foi aqui feito sobre a organização de seminários com uma frequência que nos permita ter a reflexão que sairá dos temas que aqui irão ser abordados. Essa questão é muito importante para todos nós, a comunidade que se preocupa com as questões de Educação, e é importante porque visa refletir profundamente sobre as questões de política educativa. Esse é um contributo que o Conselho Nacional de Educação começou a dar e que, aqui anunciado também pela regularidade, pode ser muito importante para a condução das políticas educativas em Portugal.

Congratulamo-nos com a temática deste seminário pois, para nós, a importância da qualidade das aprendizagens dos alunos é o que tem que orientar a ação política educativa. É, para nós, muito importante essa qualidade, porque é através da qualidade das aprendizagens que podemos propiciar aos nossos alunos percursos escolares longos, percursos escolares sustentáveis, percursos escolares duradouros e uma melhor inserção quer no ensino superior quer no mundo do trabalho. É, exatamente, pela qualidade das aprendizagens e olhar para a qualidade das aprendizagens, para nós, com determinação e sem baixar a exigência e o rigor. Porque a exigência e o rigor não fragilizam aqueles que possam vir

¹ Secretário de Estado do Ensino e da Administração Escolar.

de meios socioeconômicos mais desfavorecidos, mas, sim, tornam mais consistentes os objetivos que têm que ter em mente para seguir o mais longe possível os seus percursos escolares.

Para isso, temos que ter em conta vários fatores nessa sustentabilidade e não apenas um dos fatores, como hoje aqui vai ser analisado como primeiro fator, mas olhar de uma forma global. Desde logo que os percursos longos dependem de todos os intervenientes, dependem das políticas educativas e, conseqüentemente, dependem do Governo, através do Ministério da Educação e Ciência. É preciso saber o que se pretende para tornar as aprendizagens longas, duradouras, sustentáveis. É preciso saber se nós acreditamos que é com o reforço dos saberes fundamentais que outras aprendizagens se podem tornar mais duradouras. É preciso sabermos se dentro dos programas devem ser estabelecidas metas e se elas nos dão as indicações de como nos deveremos posicionar para que os alunos e os professores saibam, perfeitamente, aquilo que é essencial para a transição dos nossos alunos em cada ano de escolaridade. É preciso sabermos que têm que existir percursos diferenciados em função do tipo de alunos que nós temos. É preciso saber se a introdução de exames e de provas finais de ciclo são contributos para esses percursos longos e duradouros porque nos ajudam a fazer pontos de situação e atuar mais precocemente sobre conhecimentos que podem não estar a ser estruturados como o que se desejava. Portanto, essas são questões de política educativa e são questões sobre as quais devemos refletir — repercussão sobre a qualidade das aprendizagens e o percurso que se pretende dos nossos jovens.

Mas esse é um dos pontos, também é necessário olhar para o papel que têm as autarquias, porque são elas que estão mais próximas do contexto em que se desenvolvem as ações nas nossas escolas e que podem contribuir para melhores resultados e mais integração dos nossos alunos no contexto em que é desenvolvida essa ação. Também é preciso olhar para os diretores e para os professores e é preciso olhar do ponto de vista de que são eles que concebem os projetos educativos de escola e que se pretende que sejam efetivamente projetos educativos que estejam

adaptados ao contexto em que cada escola desenvolve a sua ação. Olhar para os professores e perceber que são o elemento central da aprendizagem dos alunos e valorizar essa sua mesma função. É preciso também ter em conta os encarregados de educação e a forma como valorizam socialmente a Educação. É, por isso, que existem vários fatores que contribuem para a qualidade das aprendizagens e que contribuem também para que os percursos sejam cada vez mais longos por parte dos nossos jovens, nas nossas escolas, com todas as vertentes que aqui acabaram de ser referidas.

O valor social da Educação que as famílias podem dar e reforçar aos nossos jovens é determinante, mas também a forma como os nossos jovens olham para a sua própria formação. É preciso que eles tenham consciência e é preciso que cada um dos outros atores lhes mostre que a sua vontade em aprender é determinante e não depende apenas das condições que lhes são oferecidas mas também da sua vontade, não depende apenas dos seus professores mas da sua vontade em tornar os seus percursos longos e sustentáveis.

É por isso que, para nós, é muito importante que este tema seja hoje aqui debatido — A avaliação externa e a qualidade das aprendizagens — porque a qualidade das aprendizagens tem que estar acima de tudo o que é produzido pelo sistema educativo e a avaliação externa, que tem aqui o seu enfoque nos exames ou nas provas de avaliação de final de ciclo, é uma avaliação externa que não está descontextualizada porque nós devemos olhar para a avaliação do sistema educativo como um todo integrado. É importante que façamos a nossa análise sobre a avaliação das políticas educativas, é muito importante que se tenha uma avaliação externa das escolas — como sabem, tem agora um novo modelo, desde há dois anos, a avaliação externa das escolas — avaliação externa essa que tem a sua repercussão e tem em consideração os resultados que os alunos obtêm, que faz incidir também essa avaliação externa das escolas sobre a avaliação das próprias direções dos estabelecimentos, é uma avaliação externa que tem que ter em conta também a avaliação dos professores e, neste momento aqui para reflexão, a avaliação externa que os exames e as

provas podem trazer ao sistema educativo e à orientação e ao enfoque que possa ter que ser dado a algumas das suas vertentes e à forma como está a ser feita.

Tudo isto, reforço, é algo que faz com que possamos investir na avaliação externa de uma forma integrada: políticas educativas, escolas, direções, professores, currículos, com o estabelecimento das metas, e alunos. Olhar de uma forma integrada e perceber a importância que isso pode ter, sem baixar a exigência e o rigor, o que é absolutamente necessário para que haja qualidade nas aprendizagens, mas haja qualidade nas aprendizagens para todos e não apenas para aqueles que já obtinham qualidade nas aprendizagens. Se nós baixarmos o rigor e a exigência, estamos a afastar aqueles que menos valor social dão à Educação de percursos com qualidade, de percursos sustentáveis, de percursos que se pretende que atendam às características de cada um e em que existam opções, opções diferenciadas e não obrigações, opções diferenciadas para que cada um possa, em liberdade, fazer a sua escolha, uma escolha que se pretende que traga qualidade a cada um e que não deixe ficar pelo caminho no sistema educativo, sem qualidade nas aprendizagens e logo sem percursos sustentáveis e deixando alguns deles, por falta de exigência e de rigor, sem poderem ter percursos que possam dar sequência quer no ensino superior quer no mundo empresarial.

É por tudo isso que eu queria agradecer e louvar as iniciativas que o Conselho Nacional de Educação tem levado a cabo e aquilo que perspectiva fazer. Todos nós estamos preocupados com as questões da aprendizagem dos nossos alunos e todos nós sabemos a importância que têm esses percursos e a forma como devemos construir a autonomia das escolas para acolher esses percursos e poder contextualizá-los. Cada escola serve determinados alunos e, face à autonomia que está a ser consagrada nos estabelecimentos de ensino, deve ter a liberdade de poder contextualizar aquilo que pretende que sejam os percursos dos seus alunos. O caminho da autonomia também ajudará à qualidade das aprendizagens. Se nós estivéssemos a falar há dez anos, parecia-nos quase impossível virmos a ter escolas que podem olhar para o seu currículo e construí-lo, a

flexibilização curricular era algo de distante. Isso não se deve apenas a uma decisão política, deve-se a uma necessidade que foi sempre sentida em todas as escolas e faz com que hoje se possa ter um conjunto de escolas, que se pretende que seja cada vez mais alargado, que passa a ter um nível de autonomia em que pode influenciar a qualidade das aprendizagens porque olha para o contexto em que desenvolve a sua ação educativa e pode perfeitamente adequá-lo aos seus destinatários. Isso são passos importantes na qualidade da aprendizagem.

A todos aqueles que hoje aqui vão partilhar as suas reflexões, a todos aqueles que hoje aqui vão participar nos debates queria agradecer e agradecer também o tema que foi escolhido pelo Conselho Nacional de Educação relativamente à avaliação externa e à qualidade das aprendizagens, deixando uma palavra de reconhecimento e terminar como comecei, agradecendo especialmente ao Presidente do Conselho Nacional de Educação, pela forma como nos tem feito chegar as reflexões do Conselho para que possamos, sobre elas, olhar para os percursos que estão delineados, em termos de políticas educativas, e poder tomar as decisões que possam permitir que os nossos jovens cheguem cada vez mais longe com qualidade nas aprendizagens. Para todos vós, que vão dedicar o dia a esta reflexão, o nosso muito obrigado.

Conferência
AVALIAÇÕES EXTERNAS E MELHORIA DAS APRENDIZAGENS
DOS ALUNOS: QUESTÕES CRÍTICAS DE UMA RELAÇÃO
(IM)POSSÍVEL

Presidente da Mesa – Fernando Adão da Fonseca

Domingos Fernandes¹

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 setores 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.

¹ Instituto de Educação – Universidade de Lisboa.

Seria igualmente ingênuo 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 próprias, 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 era suposto que aprendessem. 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 Benjamin 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 perspetivas 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 perspetivas ideológicas, filosóficas e sociais dos respetivos 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 dicotómicas. 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 n.º 98-A/92, de 20 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 n.º 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 abriram 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 n.º 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 (3º ou 4º anos de escolaridade) e de 13 anos (7º ou 8º 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 5º ano de escolaridade quando a grande maioria dos jovens da mesma idade dos outros países europeus participantes frequentavam o 10º ou o 11º 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 de a 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) planear 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 setores 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. Conseqüentemente, 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 perspetivas 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 de as 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 aprendam (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.

Painel
AVALIAÇÃO EXTERNA E MELHORIA DAS APRENDIZAGENS:
UMA RELAÇÃO COMPATÍVEL?

Presidente da Mesa – Jorge Ascensão

A temática em debate neste evento do CNE e a sua pertinência

Como tem sido amplamente divulgado na imprensa diária, não existe uma uniformidade de posições em relação à existência de exames nacionais, no quadro de uma avaliação externa, nem sobre os seus efeitos ao nível da progressão de alunos no sistema escolar. Neste sentido, reputo de importante o facto de o Conselho Nacional da Educação (CNE) ter elegido para este evento debater a relação entre avaliação externa, numa lógica sumativa, certificativa e comparativa, e as aprendizagens dos alunos. Espero deste debate uma reflexão que contribua para se encontrarem caminhos que permitam melhorar a aprendizagem dos alunos.

Conceções que me orientam na posição que sustento sobre a relação avaliação externa e melhoria da aprendizagem

Começando por apresentar a conceção que me orienta em relação à avaliação em geral, e portanto também em relação à avaliação da aprendizagem, refiro que me situo no grupo dos que consideram que a avaliação é parte integrante do currículo e dos processos do seu desenvolvimento, ou seja, que existe uma íntima relação entre ensino-aprendizagem-avaliação. Nesta posição considero que não se pode pensar o ensino sem se pensar na aprendizagem que ele promove mas também não se pode pensar o ensino-aprendizagem sem se pensar na sua relação com a avaliação. Por isso, atribuo enorme importância à dimensão diagnóstica da avaliação, pois considero que ela é essencial para o delineamento das etapas de qualquer processo de formação pelo apoio que confere às decisões relativas aos modos de concretizar o ensino e de estimular a aprendizagem.

A par desta ideia, também reconheço a importância da avaliação sumativa, nomeadamente como súmula de aprendizagens e de sínteses que

¹ Centro de Investigação e Intervenção Educativas – Universidade do Porto.

favorecem a articulação de conhecimentos e lhes conferem novos sentidos.

Por outro lado, aceitando que a aprendizagem pressupõe a existência de um ponto de partida, um percurso e um ponto de chegada, a avaliação, neste meu entendimento, tem um papel crucial, principalmente nas duas primeiras etapas, pela informação que fornece a professores e a alunos sobre o que já é sabido, mas também sobre o quê e onde existem dificuldades, e pela reflexão que pode suscitar na tomada de decisões quanto aos processos seguintes. Como em outro lugar sustentei (Leite, 2002), a avaliação diagnóstica é o ponto de partida para a concepção e o desenvolvimento de qualquer projeto curricular pois ela fornece elementos sobre os conhecimentos que os alunos já possuem em assuntos próximos do que vai ser ensinado, sobre características do contexto onde vai decorrer a aprendizagem e sobre competências necessárias a que essas novas aprendizagens ocorram. É a partir deste diagnóstico que as boas práticas se estruturam pois nele encontram os alicerces para estratégias curriculares que permitam a aprendizagem em espiral, isto é, onde os conhecimentos e as competências se vão continuamente ampliando. Como é evidente, se o diagnóstico for realizado ao nível do macro sistema, isto é, pelos órgãos da tutela, nesta minha concepção de avaliação ele teria também esse efeito macro, isto é, serviria para que a tutela, a partir da informação obtida, delineasse formas de ação que potenciasses os bons resultados alcançados e apoiasses os que se revelaram indesejados.

Mas continuando com esta ideia das três fases da organização dos processos de aprendizagem (ponto de partida, percurso, ponto de chegada), no caso da etapa final, isto é, na situação que nesta mesa redonda está a ser objeto de análise e que tem a ver com os testes aplicados no fim de um longo período de ensino, como é o caso dos exames ou das provas nacionais, a informação recolhida, mesmo que forneça alguns elementos de reflexão sobre o que foi conseguido durante o percurso de formação, serve essencialmente para constatar a distância face ao desejado. Ou seja, nesta etapa final a função da avaliação é privilegiadamente a de certificar e comprovar a aquisição de um conjunto

de conhecimentos e capacidades de um determinado ciclo de ensino, de selecionar quem os adquiriu e os situar numa escala hierárquica. Isto é, a função deste procedimento avaliativo não é essencialmente pedagógica, embora, pela informação que também fornece, possa e deva contribuir para tomadas de decisão futuras, elas sim com uma intencionalidade formativa.

Sendo esta a minha conceção relativa à avaliação, e afirmando desde já que vejo como necessária uma coabitação entre a avaliação de carácter formativo e a de carácter sumativo, por que considero que as provas de avaliação nacional balançam entre os efeitos menos positivos e os que podem ter reflexos positivos?

Explicitando uma posição face às provas de avaliação nacional das aprendizagens dos alunos

No sistema educativo português, e não só (Eurydice, 2010), a ideia de que a existência de exames ou de provas nacionais tem efeitos positivos na qualidade das aprendizagens tem estado presente nos argumentos que justificam a sua introdução mesmo na escolaridade obrigatória, isto é, no setor do sistema educativo que, claramente, não tem entre as suas funções a seleção.

Tendo sido abolidos os exames nacionais na escolaridade obrigatória na sequência da agenda resultante de abril de 1974, eles foram retomados na dimensão de uma avaliação aferida no princípio deste século (em 2000/2001 a título experimental). A justificação para esta avaliação aferida foi conhecer a qualidade das aprendizagens que os alunos estavam a fazer, mas também a adequação dos programas e das práticas instituídas pelas políticas educacionais. Ou seja, teve uma intenção de diagnóstico que apoiasse a introdução de alterações capazes de gerarem melhorias no currículo nacional e nas aprendizagens dos alunos. Isto é, a avaliação aferida foi justificada com uma intenção eminentemente formativa. Ainda me lembro do receio, mas simultaneamente confiança, do então Diretor da Direção Geral do Ensino Básico que, sobre as provas de aferição que foram instituídas a título experimental neste princípio de século,

considerava que elas, se bem construídas, poderiam ser indutoras de uma organização dos processos de ensino-aprendizagem mais criativos e promotores de melhores aprendizagens

Mas voltando à justificação do sentido que foi atribuído às provas nacionais de avaliação aferida, o Decreto-Lei n.º 6/2001 que reorganizou o currículo da educação básica na transição de séculos afirmou:

“As provas nacionais de aferição constituem um dos instrumentos de avaliação do desenvolvimento do currículo nacional e destinam-se a fornecer informação relevante aos professores, às escolas e à administração educativa, não produzindo efeitos na progressão escolar dos alunos” (art.º 17.º).

No caso do ensino secundário, o Decreto-Lei n.º 7/2001, também dessa altura, referiu dois tipos de avaliação sumativa (a interna e a externa) determinando que a

“A avaliação sumativa externa compreende a realização de exames nacionais no 12.º ano, incidindo sobre as aprendizagens essenciais e estruturantes correspondentes aos anos de escolaridade em que a disciplina é lecionada, nos seguintes termos: a) Na disciplina de Língua Portuguesa, em todos os cursos gerais e tecnológicos; b) Na disciplina trienal da componente de formação específica e, ainda, numa das disciplinas de opção, nos cursos gerais; c) Numa disciplina trienal da componente de formação científico-tecnológica, a definir para cada curso, nos cursos tecnológicos” (art. 11.º).

Portanto, no caso da escolaridade obrigatória, em Portugal, esta foi a forma de reintroduzir a avaliação externa das aprendizagens. No entanto, a finalidade que as orientou, pelo menos ao nível do discurso legal, era distinta dos tradicionais exames nacionais. A própria Eurydice (Rede de Informação sobre a Educação na Europa) referiu-se a estas provas como um novo método de avaliação para os 4.º e 6.º anos de escolaridade introduzidas em Portugal e que serviam de instrumento de supervisão das escolas e do sistema educativo.

Claro que quando se analisa esta questão torna-se evidente que a reintrodução de processos de avaliação externa em Portugal foi, e tem sido, fortemente influenciada por políticas internacionais que a referem como meio de promoção da qualidade educacional. Exemplo disso são os estudos PISA (2000-2009) que afirmam que ela constitui um dos fatores para que um sistema de ensino se desenvolva e progrida.

Mas será que estas provas nacionais apenas têm esta marca positiva?

Em Portugal, depois do recurso a esta lógica da avaliação externa na forma de avaliação aferida, para o caso do ensino básico, ela passou a ser assumida na forma de exame nacional, no caso do 3.º ciclo em 2006, e só para as disciplinas de Português e Matemática (aliás, tal como acontecia nas provas de avaliação aferida). O facto de a avaliação se concentrar nestas duas disciplinas, e embora constitua uma situação semelhante ao que se passa em muitos outros países, impõe que se interroge, pelo menos, porquê só estas disciplinas. Se a avaliação externa influencia a qualidade das aprendizagens, que razão justifica avaliar-se apenas Português e Matemática?

Embora sejam várias as explicações que podem e têm sido apresentadas para justificar a concentração dos exames nestas duas disciplinas não se pode ser inocente em relação a esta questão, ou seja, é evidente o peso dos relatórios internacionais, de que o PISA é um dos exemplos, nesta decisão política. Recorde-se que o PISA (Programme for International Students Assessment), realizado pela OCDE, avalia conhecimentos demonstrados por alunos de 15 e 16 anos de diferentes países relativamente às capacidades de leitura e de interpretação, e de conhecimentos de matemática e de ciências.

O que estou a afirmar é que é no quadro deste movimento internacional que tem ocorrido o recurso a exames ou provas nacionais em Portugal. Em 2012 foram introduzidas as provas nacionais no 6.º ano nas áreas de Matemática e Língua Portuguesa, que contam com 30% para a classificação (nesse ano 25%), substituindo as provas de aferição, e em

2013 foram introduzidas as provas nacionais para o 4.º ano, também em Matemática e em Português.

O relatório do GAVE desse ano (2013) refere que, decorrente dos resultados da prova, na 1.ª fase, não podiam transitar para o 2.º ciclo 628 alunos a Português e 2100 alunos a Matemática (3,6% do total dos alunos) e que, depois da medida “acompanhamento extraordinário” e da realização da 2.ª fase, a taxa de retenção baixou para 3% (nos anos anteriores tinha sido mais elevada – em 2012 era 5%). Ou seja, deste caso poder-se-ia inferir que estas provas têm como aspeto positivo desencadear ações que promovem aprendizagens.

Apesar disso, não podem ser ignoradas as posições daqueles que criticam este tipo de provas nacionais por desvalorizarem determinadas áreas de formação e por avaliarem apenas conhecimentos e dimensões cognitivas da aprendizagem quando a educação básica, nas suas funções, tem também uma forte intervenção em aprendizagens sociais.

Reconhecendo como legítimo este tipo de argumentos, reconheço no entanto também que as provas nacionais podem ter o efeito positivo de corresponsabilizar escolas, professores e até famílias na necessidade de encontrarem meios que promovam o sucesso de todos os alunos. Devo referir que é o contacto que tenho tido com escolas e professores que me leva a acreditar neste efeito. Aliás, esta é a orientação que está presente no Programa dos Territórios Educativos de Intervenção Prioritária (TEIP), quando as escolas definem um plano de melhoria, estabelecem metas que a concretizem e depois prestam contas sobre o modo como as estão, ou não, a conseguir atingir. O mesmo se passa com os Contratos de Autonomia das Escolas negociados com o Ministério da Educação. O trabalho de assessoria que tenho desenvolvido com um Agrupamento de Escolas da zona Norte de Portugal, e que, estando incluída no Programa TEIP, celebrou um contrato de autonomia com o Ministério de Educação, tem-me mostrado os feitos positivos que decorrem da reflexão gerada a partir dos resultados dos exames nacionais. O conhecimento obtido através dessas provas de avaliação sumativa externa tem constituído um

elemento importante para analisar os projetos curriculares e para o delineamento de novas formas de ação.

No entanto, e apesar de reconhecer este efeito positivo, também considero que não basta definir metas, prestar contas e sujeitar os alunos a provas de avaliação nacionais para que as aprendizagens melhorem. Se isso fosse verdade, os resultados obtidos pelos alunos neste tipo de provas estariam a subir cada ano e tal não tem acontecido. Por outro lado, há que não ignorar os efeitos negativos que estas provas muitas vezes geram. Todos sabemos que, associado a esta avaliação externa, a comunicação social tem divulgado um *ranking* de escolas, com grande visibilidade social, e que tem tido como consequência que as escolas com classificações mais baixas tenham passado a ser cada vez menos escolhidas por alunos de determinados grupos sociais e mais frequentadas por crianças e jovens mais afastados da cultura escolar. Ao contrário, as escolas melhor colocadas nos *rankings* são procuradas pelas famílias que têm para os seus filhos projetos de vida que passam pelos diplomas escolares. Ou seja, está a ocorrer o Efeito Matthew, no sentido das vantagens acumuladas/cumulativas de que nos falou Robert Merton (1988) a propósito da influência de um sucesso inicial para obter novos sucessos.

Por isso, de certo modo, o que estou aqui a referir está de acordo com os argumentos daqueles que se têm pronunciado contra os exames ou provas nacionais por eles apenas servirem para legitimar diferenças resultantes da origem da população escolar e não da qualidade do serviço prestado em cada escola. Bem sei que, nos últimos anos, são tidas em conta variáveis de contexto mas, mesmo assim, este efeito secundário continua a ser visível. Embora muitas escolas e muitos professores estejam a assumir o importante papel de criarem situações e desenvolverem projetos que contrariem o efeito Matthew, anulando, ou pelo menos minimizando, as desvantagens que decorrem das situações de partida, muitas vezes esse envolvimento não se traduz na conquista de um bom lugar no *ranking*. Apesar disso, muitos destes professores conseguiram despertar, em alunos que não correspondem ao que H. Becker designou por “cliente ideal”, novas atitudes face à educação escolar, ao mesmo tempo que promoveram

o desenvolvimento linguístico e cognitivo dos alunos e lhes proporcionaram vivências positivas de cidadania.

Se penso assim, por que razão ainda reconheço que as provas nacionais têm alguns efeitos na aprendizagem, principalmente, das disciplinas sujeitas a exame nacional?

Como já indiciei, o contacto que tenho tido com escolas, quer decorrente de projetos curriculares, quer da assessoria a um agrupamento de escolas TEIP, quer ainda de avaliação externa das escolas leva-me a reconhecer que, apesar dos efeitos perversos que enunciei, existem aspetos positivos que podem decorrer das provas nacionais. O facto de os resultados obtidos pelos alunos em provas nacionais serem objeto de uma forte atenção por parte do Ministério da Educação e da sociedade em geral tem como efeito que escolas e professores façam um esforço grande na procura de processos de melhoria das aprendizagens. Há dias, uma coordenadora de departamento de um agrupamento de escolas TEIP lamentava-se por estar a reconhecer que alguns dos seus alunos do 9.º ano eram incapazes de adquirir alguns conhecimentos necessários na Matemática e que, para além disso, aparentemente, isso não os molestava. Apesar disso, quando lhe perguntei a opinião relativa à existência, ou não, de provas nacionais a resposta foi afirmativa. Reconhecia que, apesar de tudo, estas provas tinham alguns efeitos em alguns alunos, em algumas famílias e até no esforço feito pelos professores para conseguirem vencer alguns dos insucessos que ocorrem.

Claro que, a par desta opinião, o coordenador do departamento das Expressões manifestava que a atenção focalizada nestas duas disciplinas tinha como efeito uma desvalorização, por parte dos alunos e das famílias mais atentas aos percursos escolares dos seus filhos, de toda a formação que não sofre desta pressão e desta visibilidade social.

Em síntese, para terminar e sistematizar a minha posição face a este debate tenho de reconhecer que fico em cima do muro, saltando, umas vezes, para o lado das opções e ações que promovem uma formação global e, por isso, em vez de se acantonarem na preparação técnica dos alunos apenas para

serem bem-sucedidos nos exames, valorizam a importância da escola na socialização das crianças e dos jovens para uma vivência positiva da cidadania. Mas, regressando acima do muro para me manter atenta aos efeitos positivos que poderão decorrer da prestação de contas através de provas nacionais, reconheço o conhecimento que através delas por vezes se obtém, quer para sustentar balanços do sistema educacional, quer para apontar caminhos ao sistema, às escolas, aos professores e aos alunos sobre o que é preciso fazer para se melhorar a qualidade da formação que está a ocorrer nas escolas a nível nacional. No entanto, e como penso ter tornado evidente, quando estou deste lado do muro considero que a responsabilidade de usar positivamente a informação que as provas nacionais fornecem não é apenas das escolas, dos professores e dos alunos mas também da administração central para que sejam desenhadas intervenções indutoras de uma real e coletiva melhoria.

Referências

- BECKER, H. (1977). Social-class variations in the teacher-pupil relationship, in *School and Society. A sociological reader*/ Open University.
- GAVE (2013). Relatórios nacionais, in www.gave.min-edu.pt/np3/24.html
- LEITE, C. (2002). "Avaliação e projectos curriculares de escola e/ou de turma", in ME/DEB, *Avaliação das aprendizagens. Das concepções às práticas*, pp. 43-51.
- EURYDICE (2010). Exames nacionais de alunos na Europa: objectivos, organização e utilização de resultados, in eacea.ec.europa.eu/education/eurydice/documents/thematic.../109PT.pdf
- MERTON, R. K. (1988). [The Matthew Effect in Science, II: Cumulative advantage and the symbolism of intellectual property](#) (PDF). *ISIS* 79, 606-623.

AVALIAÇÃO EXTERNA, AVALIAÇÃO INTERNA E APRENDIZAGEM ESCOLAR

João A. Lopes¹

Os efeitos da avaliação externa sobre as aprendizagens são complexos e de difícil mensuração. Antes de mais devido à dificuldade de definição dos conceitos envolvidos. Se a definição de avaliação externa é relativamente simples, a definição de aprendizagens dos alunos é bem mais complexa. Ainda assim, no caso da avaliação externa há que diferenciar a avaliação externa através de exames nacionais da avaliação externa através de provas internacionais, sendo que, nos dois casos, as populações abrangidas só parcialmente coincidem (pelo menos na forma directa de realização de exames. Na forma indirecta, que resulta eventualmente da pressão social sobre os envolvidos, a sobreposição parece maior).

A definição de aprendizagens dos alunos é, como se referiu, de mais difícil definição e, conseqüentemente, de mais difícil mensuração. Porventura a forma mais simples será a de avaliar o impacto dos exames por comparação com os resultados dos exames seguintes. Contudo, isto levanta o problema de saber se a avaliação será intra-sujeito ou intra ano lectivo. No primeiro caso, o sujeito estará a ser avaliado em provas diferentes (por exemplo, no 4.º ano e no 6.º ano) e em anos diferentes, o que pode ser interessante do ponto de vista individual mas que levanta alguns problemas quando se está a considerar o conjunto dos alunos. No segundo caso (em que, por exemplo se avalia os resultados no final do 4.º ano em anos lectivos consecutivos), os sujeitos não são os mesmos, o que também levanta alguns problemas de interpretação.

As dificuldades de avaliação do impacto da avaliação externa nas aprendizagens dos alunos são ainda agravadas pelo facto de haver um conjunto significativo de variáveis mediadoras entre o preditor (os exames) e os resultados (as aprendizagens dos alunos). De entre estas

¹ Escola de Psicologia – Universidade do Minho.
O autor escreve de acordo com a antiga ortografia.

variáveis mediadoras poderemos relevar os pais, as escolas, os professores, as expectativas dos alunos, etc. Cada uma destas variáveis pode ser simultânea ou alternadamente moderadora, causal, etc., relativamente às aprendizagens mas também relativamente aos exames em si mesmos.

Seja qual for a relação e a estrutura de influência das diversas variáveis nas aprendizagens dos alunos (independentemente da forma com estas forem conceptualizadas e medidas) é possível supor que a cultura sistemática de avaliação e mensuração que se instalou há pouco mais de uma década em Portugal e que também se tem acentuado em termos internacionais, tem provocado alterações nos comportamentos de todos os envolvidos nesta questão.

Exames e publicitação dos resultados

O que parece mais relevante considerar, quando se procura explicar a alteração de comportamentos em função dos resultados da avaliação externa, é o efeito da publicitação dos resultados e não tanto a avaliação em si. Deve lembrar-se que o debate em torno do conhecimento público dos resultados dos exames do 12.º ano foi alvo de um acalorado debate há cerca de 15 anos atrás, tendo havido um forte resistência por parte da administração pública, cujos argumentos foram variando ao longo dos anos mas tiveram sempre como denominador comum a noção de que apenas ela, administração, saberia como utilizar adequadamente os dados. Ainda que de forma mais esbatida, o antagonismo de posições entre os que pensam que a informação deve ser do conhecimento público e aqueles que defendem que não, ainda hoje se mantém na sociedade portuguesa.

Precisamente, e como se disse atrás, parece ter sido a divulgação pública das notas nos exames do 12.º ano, iniciada há anos atrás, mais do que os resultados do TIMSS, PISA, etc., a desencadear um movimento organizacional (i.e. de escolas individuais) com orientação clara de comportamentos colectivos e individuais para preparação para as provas de avaliação. Muitas escolas terão, com esta divulgação, sentido necessidade de mostrar melhores resultados e de fornecer uma melhor

imagem de si ao público em geral. A generalização dos exames externos a outros ciclos de ensino terá acelerado este movimento e terá estendido a pressão para a realização nos exames às famílias e à sociedade em geral. Os “*sites*” de muitas escolas, com referências sistemáticas à progressão dos resultados nos exames nacionais, evidenciam de forma particular esta preocupação. Os “*rankings*” organizados pelos jornais e televisões acentuam esta tendência, sendo difícil que alguém lhes fique indiferente.

O confronto de muitas escolas com o desempenho de outras escolas com características diferentes tem permitido às escolas que apresentam resultados inferiores justificarem os seus piores resultados com base nas diferenças de nível socioeconómico dos alunos. Mais difícil porém se torna explicar as diferenças de resultados por comparação com escolas de níveis socioeconómicos idênticos.

Independentemente da elaboração de auto-justificações que permitam diminuir a frustração de ver um resultado desagradável publicitado num jornal, parece inquestionável que em muitas escolas houve alterações institucionais que resultaram numa focalização da acção dos professores e do trabalho dos alunos nos conhecimentos e competências avaliados em exame final externo. Em muitas escolas os alunos começaram a ser treinados para os exames finais, inclusivamente em situações de simulação literal das provas finais. A pressão sobre as direcções das escolas exercida pela publicitação dos resultados publicados nos jornais, pelos pais e pela própria administração da educação, bem como a pressão exercida pelas direcções sobre os professores e sobre os alunos, parece ter tido como efeito uma contração na preparação para os exames. De resto, a subida sucessiva dos resultados dos alunos portugueses de 15 anos no PISA revelam uma melhoria de fundo que provavelmente terá a ver com o conjunto de medidas que foram entretanto adoptadas para melhorar o desempenho dos alunos portugueses. Deste modo, parece elementar dizer que a publicitação dos resultados dos exames tem tido um efeito positivo, mesmo que indirecto, nos desempenhos no PISA e que isso representa uma melhoria nas aprendizagens dos alunos portugueses.

Dados da investigação sobre o impacto da avaliação externa

Apesar da complexidade da avaliação do impacto dos exames externos nas aprendizagens dos alunos, existem diversos estudos internacionais (e.g. Anghel, Cabrales, Sainz, & Sainz, 2012; Jorges, Schneider, & Buchel, 2003) que procuram medir este impacto e perceber se ele é relevante para os sistemas educativos e para o desempenho dos alunos.

Anghel et. al. (2012), aproveitando o facto de a região de Madrid ter sido em 2004/05 a única região de Espanha que efectuou exames nacionais do 6.º ano e publicitou os resultados, estudou os efeitos destes exames e desta publicitação nos resultados dos alunos. Utilizando uma metodologia “diff-in-diff” ao autores encontram uma diferença de 14 a 17 pontos no PISA, favorecendo os alunos da região de Madrid que se submeteram à avaliação externa. Os autores sublinham que o seu estudo confirma o que foi verificado noutros estudos mas que o seu próprio estudo tem a particularidade de os resultados do exame não terem consequência real para os alunos, pelo que o efeito advirá previsivelmente do impacto da avaliação externa nos professores e nas direcções das escolas. Os autores encontraram igualmente um possível efeito dos exames externos em provas de linguagem mas não na matemática.

Jorges, Schneider, e Buchel (2003), por seu turno, sugerem que os alunos dos estados federados alemães que realizam exames centralizados ultrapassam largamente os alunos dos outros estados mas reconhecem que apenas uma parte dessa diferença poderá ser atribuída aos exames externos. Reconhecem igualmente que identificar empiricamente uma relação causal entre realização de exames externos e aprendizagens dos alunos não é linear. De qualquer forma os autores afirmam que os seus resultados mostram que os alunos dos estados alemães com exames ultrapassam os dos seus colegas dos estados sem exames em cerca de 0,5 desvios-padrão. Os autores frisam, por outro lado, que os exames constituem uma forma relativamente barata de melhorar os resultados dos alunos, por comparação com soluções caras e de eficácia duvidosa, como é o caso da redução do número de alunos por turma.

Aspectos a acautelar na utilização dos resultados da avaliação externa

Independentemente do eventual impacto positivo da avaliação externa nas aprendizagens dos alunos, existem aspectos que poderão ser menos positivos neste processo. Antes de mais deve evitar-se transformar uma importante e legítima preocupação de avaliação e monitorização do sistema (que perpassa todo o sistema de ensino, do básico ao superior), numa obsessão psicométrica cuja validade é duvidosa em algumas áreas. Nem tudo nas aprendizagens é facilmente avaliável, havendo por isso o risco de apenas ser valorizado aquilo que é avaliável ou de toda a acção do sistema se orientar para a avaliação nessas áreas e com essa configuração.

Um outro eventual problema poderá advir da eventual hipervalorização e hipertrofia de certas áreas em detrimento de outras. Ainda que se possa considerar que a Língua Materna e a Matemática são instrumentais para outras áreas disciplinares, sobretudo no ensino básico, a exclusividade destes exames envia inevitavelmente uma mensagem quanto à valorização social de certas áreas em detrimento de outras. A possibilidade de algumas áreas definharem é real e preocupante. Poderia, a título de exemplo, referir-se a desvalorização sucessiva a que disciplinas como a Filosofia ou o Latim têm estado sujeitas.

Há ainda o risco de em alguns casos a forma como muitas escolas divulgam os seus sucessos no seus “sites” (quer no que diz respeito a resultados nos exames quer no que diz respeito à disciplina e segurança na escola) não corresponder rigorosamente à substância dos resultados. A necessidade que muitas escolas sentem em lutar por alunos (e não já em lutar pelos melhores alunos) leva a que algumas divulguem com grande ênfase os saltos de “centenas de posições” no “ranking” das escolas. Alguns desses saltos são tão extraordinários que poderão levantar dúvidas quanto aos critérios de inclusão e de exclusão de alunos, dos exames. Alguns países têm sido acusados de excluir das avaliações internacionais alunos com dificuldades, fazendo assim subir artificialmente os resultados. Embora isso seja eventualmente mais difícil de fazer a nível de escola, será necessário em todo o caso averiguar se os progressos inesperados de

certas escolas se devem mesmo a alterações eficazes da estratégia ou a artefactos ilegítimos. Num quadro de extrema competição pelos alunos por vezes surge a tentação de alcançar resultados a todo o custo e a que o sistema educativo deve estar atento.

Um outro aspecto a que convirá dedicar atenção é ao custo dos exames. Neste particular, deve realçar-se que actualmente o número de exames organizados pelo IAVE (Instituto de Avaliação Educacional), entre provas nacionais e internacionais, é enorme. A logística destes exames é extremamente complexa, sendo os meios disponíveis muito escassos. O esforço anual do IAVE é importante e não pára de aumentar de ano para ano. É possível que do ponto de vista do custo humano se esteja à beira dos limites do exigível. Do ponto de vista financeiro, não deixa de impressionar o enorme número de páginas de teste entregue a cada aluno, o que sugere que o país não tem qualquer dificuldade em custear as provas.

É evidentemente possível encolher fortemente estes custos sem prejuízo da validade das provas. Se o sistema for demasiado caro para o retorno que se obtém, algo terá que, neste particular, ser repensado. Note-se que os exames são, segundo alguns autores (e.g. Anghel et. al., 2012) uma forma muito mais barata de melhorar os resultados, do que reduzir o número de alunos por turma. Nesta perspectiva é conveniente atender aos custos financeiros e humanos dos exames para perceber se são mesmo a forma mais barata de melhorar os resultados. E mesmo que o sejam, é importante perceber se não custam demasiado dinheiro.

Um outro ponto de grande relevo é o da articulação da avaliação externa com a avaliação interna e, mais ainda, com o desenvolvimento de estratégias que permitam efectivamente melhorar as aprendizagens dos alunos. É importante notar que a avaliação externa se limita a constatar a situação do sujeito num determinado momento do seu percurso escolar mas não permite a sua correcção atempada. Esta correcção só pode decorrer da instituição de um eficaz sistema de monitorização interna que possibilite a intervenção imediata sobre os problemas de aprendizagem detectados. Neste particular é importante notar que a avaliação no 4.º ano

pode servir para as escolas reflectirem sobre os seus resultados mas já não permitirá a correcção de percursos individuais. Esta correcção só pode decorrer do processo de avaliação interna realizada pelo professor ou por escolas individuais. Este ponto, que parece de importância menor, ganha um enorme relevo se se atentar no facto de a probabilidade de recuperação dos alunos com dificuldades no 5.º ano (após os exames) estar praticamente votada ao fracasso. Modelos de avaliação/intervenção como a RTI (Response to Intervention) têm o potencial de preencher esta lacuna, apesar de actualmente serem praticamente desconhecidos em Portugal.

Um último e importante ponto deve ser realçado: o esforço que o país está a fazer com a organização dos exames externos (nacionais e internacionais), pode estar a ser prejudicado, minado, ou até mesmo tornado pouco produtivo, pela percepção de aumento da indisciplina nas salas de aula manifestado por muito professores. Demasiados professores parecem percepcionar a sua profissão como uma luta continuada com alunos, pais, e até governos, tendo como consequência um desgaste diário significativo que prejudica o ensino e a aprendizagem. Nestas circunstâncias o impacto dos exames sobre as aprendizagens dos alunos será certamente menor do que seria esperado caso os professores não tivessem que desperdiçar com a indisciplina uma preciosa parte da energia que deveriam utilizar com o ensino, fundamento essencial da acção do professor.

Conclusão

Parece ser consensual que a avaliação externa tem impacto sobre os resultados escolares, ainda que este impacto seja dificilmente mensurável. Mais do que os exames em si, parece ser a divulgação pública dos resultados a alterar significativamente os comportamentos individuais e colectivos nas escolas sob avaliação. A resistência inicial à instituição dos exames parece ser hoje bem menor e mais mitigada, o mesmo sucedendo quanto à publicitação dos resultados. Deste modo, parece razoável considerar que há hoje um maior reconhecimento do impacto positivo dos exames nas aprendizagens. Subsistem porém argumentos não

despiciendos quanto à orientação porventura excessiva do ensino para a realização dos exames e/ou quanto à possibilidade de as disciplinas não-sujeitas a exame poderem ser subalternizadas (argumento que acaba por reconhecer o impacto dos exames). Por outro lado, ainda que os exames se constituam como reguladores da acção dos professores e das escolas, não se poderá esperar deles que originem mais do que aquilo que podem originar.

Por fim haverá que ponderar os custos dos exames enquanto estratégia de melhoria dos resultados dos alunos, mantendo-os num patamar que não seja excessivo, nem do ponto de vista dos recursos humanos nem do ponto de vista dos recursos financeiros.

Referências

- ANGHEL, B., Cabrales, A., Sainz, J., & Sainz, J. (2012). Publicizing the results of standardized external tests: Does it have an effect on school outcomes? Departamento de Economía - Universidad Carlos III de Madrid (Ed.), *Working paper - Economic Series*.
- HENDRIK JÜRGES & Kerstin Schneider & Felix Büchel, 2005. The Effect Of Central Exit Examinations On Student Achievement: Quasi-Experimental Evidence From TIMSS Germany. *Journal of the European Economic Association*, vol. 3(5), pp.1134-1155.

Francisco Vargas¹

Evaluar suele ser un proceso muy complejo que implica una variedad de factores y que tiene consecuencias en diversos ámbitos de la vida de las personas. Todos los días, al interactuar con el mundo que nos rodea, somos evaluados desde diferentes perspectivas por personas con diferentes parámetros y que nos observan y califican para diferentes objetivos. De allí que la evaluación generalmente sea percibida como el aspecto más importante de muchos procesos y esté asociada con situaciones de estrés y con consecuencias, incluso llegando a ser usada como castigo.

Sin embargo desde la perspectiva educativa, la evaluación no debe ser el componente primordial del proceso enseñanza-aprendizaje sino un componente más de dicho proceso. En este sentido la evaluación debe estar alineada con el perfil de alumno que los colegios desean: Si un colegio plantea que su perfil de alumno se basa en la “adquisición” de una gran cantidad de contenido que el alumno debe recordar y reproducir en un examen, entonces sus evaluaciones serán típicos exámenes que miden básicamente la capacidad que tiene el alumno para recordar hechos y datos concretos. Por otro lado, si el perfil de alumno se basa en el desarrollo de habilidades académicas concretas, en el desarrollo de conceptos que trascienden los límites de las disciplinas y en la capacidad de transferencia lejana y cercana de conocimientos y habilidades a situaciones complejas, entonces las evaluaciones serán una amplia variedad de instrumentos que incluyen exámenes, estudio de casos, actividades experimentales, investigaciones matemáticas, obras de arte, paginas web, actividades interdisciplinarias y tantos otros instrumentos como la creatividad del docente y la complejidad de las realidades elegidas a estudiar sean capaces de generar.

En este contexto, la discusión cambia de perspectiva y ya no se centra entre la conveniencia o no de evaluaciones externas sino en cuales son los

¹ Head of Curriculum & Vice-Principal Oeiras International School.

contenidos y las habilidades que las evaluaciones tanto internas como externas deben medir.

Al revisar la literatura disponible relacionada con la habilidad de pensamiento y la taxonomía que clasifica y ordena estas habilidades, nos encontramos con cambios importantes desarrollados en la última década, por ejemplo, en la taxonomía de Bloom cuya versión revisada por Anderson, se resume en el siguiente diagrama.

<i>Habilidades de pensamiento de orden superior.</i>
CREAR
EVALUAR
ANALIZAR
APLICAR
COMPRENDER
RECORDAR
<i>Habilidades de pensamiento de orden inferior.</i>



Desde la perspectiva de esta taxonomía y del perfil de alumnos que un colegio o un sistema educativo desea obtener al final del proceso, la discusión entonces se centra en que tipo de evaluaciones externas son adecuada para medir los conocimientos y las habilidades de los alumnos y no si las evaluaciones externas son adecuadas o no. Además hay una ventaja evidente cuando se aplican evaluaciones externas: Estas evaluaciones son un estándar externo al proceso que no está influenciado por emociones u otros factores, por lo tanto sus resultados pueden ser considerados objetivos.

En resumen, la implementación de evaluaciones externas que consideren diferentes estilos de aprendizaje, inteligencias múltiples y habilidades superiores del pensamiento seria una muy buena herramienta para medir el

logro de un proceso enseñanza-aprendizaje diseñado y desarrollado bajo los mismos preceptos.

Valter Lemos¹

Porque é que é necessária a avaliação externa? A resposta está numa frase do prémio Nobel François Jacob, que nos diz que “a familiaridade provoca a cegueira” e portanto nós tendemos a deixar de ver os pormenores naquilo que nos é familiar e isso leva-nos, muitas vezes, a pedir, ou a ter necessidade, numa linguagem de investigação, de criar um afastamento, para vermos alguns aspetos que não vemos quando as questões nos são demasiado familiares. É, do meu ponto de vista, uma das razões para que a avaliação externa seja uma necessidade social porque, como já foi referido por Domingos Fernandes, acima de tudo, estamos a falar de um processo social.

Relativamente às questões da educação, às aprendizagens, a pergunta de fundo, e eu creio que é a pergunta que está subjacente a este seminário e a várias das intervenções que foram feitas, é: Mas deve ou não existir avaliação externa das aprendizagens? Ou seja, quando estamos a falar da questão das aprendizagens dos alunos, a primeira pergunta é: Deve ou não deve existir? Deve ou não deve ter lugar? Há, realmente, duas razões essenciais para a existência de avaliação externa das aprendizagens. A primeira razão é a confiança social, pelo que disse anteriormente, ou seja, como a familiaridade provoca a cegueira o que acontece é que há alguma desconfiança dos aspetos resultantes do juízo sobre alguma coisa que nos é familiar. A avaliação, enquanto processo social tem funções como a certificação e a seleção, que, para garantir a confiança social, exigem a dimensão externa da avaliação.

Por outro lado, a educação, hoje, tornou-se uma das maiores indústrias do mundo moderno, as escolas (e os centros comerciais...) são os únicos sítios onde há concentrações permanentes de multidões e têm dimensões sociais apreciáveis no sentido quantitativo e qualitativo. A educação entrou completamente no mundo das pessoas, de todas as pessoas, e não deixa ninguém de fora — as escolaridades são obrigatórias, cada vez mais

¹ Escola Superior de Educação – Instituto Politécnico de Castelo Branco.

longas — esta máquina é cada vez maior e cada vez mais complexa e isso exige, naturalmente, uma regulação política cada vez mais difícil, cada vez mais exigente, precisamente porque a sua natureza é cada vez maior e mais complexa. Essa regulação política exige distanciamento e a necessidade de ter indicadores que lhe permitam tomar decisões e que possam ser cruzados com a tal informação que vem da maior familiaridade, ou seja, a informação que vem diretamente das escolas, dos professores e dos alunos.

Mas, isto coloca um problema que é o seguinte: Quando referimos a avaliação externa do que estamos a falar em concreto? Quando olhamos em termos imediatos diremos que a avaliação que não é da responsabilidade direta dos professores ou das escolas é aquilo a que chamamos avaliação externa. Mas a complexidade desta máquina que são os sistemas educativos modernos originou o desenvolvimento de uma panóplia enorme de instrumentos de avaliação. Os instrumentos vão-se criando, vão-se diversificando, cada vez há mais instrumentos porque também há mais complexidades, mais objetivos, mais atores e isso obriga à sua multiplicação. Há quem refira, no entanto, que tecnicamente os instrumentos são sempre os mesmos, mas na verdade não são, pois os instrumentos não são só a sua definição técnica, são também, o seu uso, a sua utilização, as suas consequências, os seus efeitos. Os instrumentos também são isso. Os instrumentos não são exclusivamente os testes em concreto, são o uso que lhes é dado, a forma como são concebidos, a forma com são aplicados, a forma como é utilizada a informação que deles provém, etc.

Na verdade eu diria que existem três tipos essenciais de instrumentos muito distintos, provavelmente menos distintos no seu aspeto, mas absolutamente distintos no seu objetivo, na sua finalidade, ou seja, no que fazemos com os seus resultados e esses instrumentos têm tempos de vida completamente diferentes na história social e na história da política educativa. Eu distingo esses três instrumentos por três diferenças: os exames com um objetivo de fazer uma certificação social (não é uma certificação escolar, eu insisto, é uma certificação social), as provas

aferidas que pretendem monitorizar os programas, os atores, os contextos, etc., (e por isso podem não se aplicar a todos); e os testes internacionais que também têm a mesma natureza, mas que nasceram com um objetivo muito diferente, o de comparar políticas educativas (não se pode ignorar isto, porque o objetivo não é só monitorizar os sistemas de ensino, mas, através dessa informação, fazer também a comparação e a avaliação das políticas).

Uma nota sobre os testes internacionais e, nomeadamente, o PISA, porque este ganhou uma dimensão, hoje, que eu diria que, praticamente, não há nenhum responsável de política educativa no mundo, pelo menos no mundo desenvolvido, que possa olhar de forma menos atenta para o PISA. Ganhou o espaço de conseguir transformar-se realmente num instrumento de verdadeira monitorização das políticas e hoje esse objetivo está presente, de forma clara e expressa, nos seus objetivos. Deverá ser difícil encontrar na Europa um discurso político sobre educação ou um programa de um governo que de alguma forma não vá buscar aspetos de fundamentação ao que o PISA nos diz.

Esse impacto tem sido enorme em termos de política educativa. Recordo que desde os primeiros resultados do PISA a que Grek (2009) chamou o “*PISA shock*” quando saíram os primeiros resultados na Alemanha e o “*PISA surprise*” na Finlândia, precisamente pela diferença que existia nos resultados obtidos e aquilo que eram as expectativas que esses países tinham relativamente aos seus próprios resultados e as expectativas que outros países tinham relativamente a esses resultados. Um estudo recente (Breakspear, 2012) que envolve 37 países mostra que hoje o PISA se transformou realmente num *standard* internacional de *performance*, não dos alunos, mas dos sistemas escolares dos vários países. Aparentemente, os resultados dos alunos já são um instrumento do verdadeiro objetivo do PISA.

Mas afinal, na avaliação externa em Portugal, qual é o problema que verdadeiramente não está resolvido e que nós oscilamos para ver se conseguimos ou não resolver? É o problema da existência dos exames e/ou da existência das provas aferidas como instrumento central da

avaliação externa. E a minha convicção é de que não é possível fazer as duas coisas ao mesmo tempo e o que nós tentamos fazer é mascarar, isto é, vamos tentando disfarçar, tentando usar uma coisa para as finalidades da outra, os exames com a finalidade de monitorizar o sistema e, eventualmente o contrário, que é usar as provas aferidas com a finalidade de certificar os resultados dos alunos. Há uma diferença clara nas finalidades, no referencial, no critério e na forma de fazer o feedback. Basta ler os relatórios do GAVE sobre as provas aferidas e ler os relatórios sobre os exames e ver a diferença como são organizados os dois tipos de relatórios, para perceber que o feedback dado por cada instrumento é bastante diferente e, portanto, o problema é: com a avaliação externa pretendemos fazer a certificação social (os diplomas escolares têm essa função), ou pretendemos utilizar essa informação para a melhoria das aprendizagens? (Esta última é a pergunta que nos foi colocada neste seminário)

Gostaria de fazer uma nota final sobre a questão porque, do meu ponto de vista, esta deriva de transformar as provas aferidas de 4.º ano e 6.º ano em exames é uma deriva perigosa. A primeira questão é, se são exames, o que nós sabemos dos exames é que eles servem para certificar, mas não servem para o resto, como aliás já foi explicado pela Carlinda Leite, e muito bem, quando disse que, se os exames nos dessem esse feedback de melhoria, há tantos anos que há exames, os resultados estariam sempre a melhorar.

Os exames existem há mais anos que todas as outras formas de avaliação externa, tanto em Portugal com nos outros países, e Portugal é um país que já passou pela situação de ter exames e de não ter exames e de ter provas aferidas e de não ter. Portugal tem essa experiência, que é uma experiência bastante interessante e diferente de grande parte dos países e eu deixo uma nota para reflexão: a minha convicção é que a avaliação externa é extraordinariamente importante para a melhoria das aprendizagens mas não exames com o formato que lhe temos dado. Os exames extinguem-se com o ato de certificação, são provas gerais de domínio e não contêm a parte seguinte que é a que verdadeiramente conta para a melhoria das

aprendizagens, ou seja, a informação que permite o feedback às escolas e aos professores sobre as competências específicas dos alunos.

Para a avaliação externa ter efeitos na melhoria das aprendizagens é indispensável que os seus resultados sejam sistematicamente devolvidos às escolas, aos professores e aos alunos e tratados de forma analítica em função das competências avaliadas e da sua relação com o tempo e os atores.

A avaliação externa, como qualquer avaliação sumativa, relaciona-se com a melhoria das aprendizagens quando... passa a formativa, isto é, quando nos permite passar do juízo à interpretação e à compreensão.

Por outro lado, para aqueles que acham que os exames são condição promotora da qualidade do ensino e da melhoria das aprendizagens, eu fiz alguns cálculos sobre os anos em que não houve exames em Portugal e cheguei à conclusão que os portugueses que têm 35 ou 36 anos e 53 ou 54 anos não fizeram qualquer tipo de exame até ao final do ensino secundário (fizeram, quanto muito, as provas de acesso ao ensino superior que não é relevante para o que estamos a discutir aqui). Assim a minha pergunta é, para aqueles que acreditam que os exames são um instrumento indispensável à garantia da qualidade das aprendizagens (não à garantia da qualidade da certificação social), e a pergunta é se acham que, no caso dos portugueses entre os 35 e os 50 anos, estamos perante uma geração de impreparados ou de falhados. Eu estou convictamente pensando que não, e que, bem pelo contrário, estamos perante uma geração muito bem preparada e alguns dos melhores cientistas portugueses são desta geração que não fez nenhum exame no ensino básico e no ensino secundário.

A minha convicção é essa, mas a pergunta fica para a reflexão de todos.

Painel
AVALIAÇÃO EXTERNA E
PRÁTICAS DOCENTES

Presidente da Mesa – Diogo de Lucena

VESTIR A CAMISOLA DO EXAME E A DA PRÁTICA LETIVA

Aldina Silveira Lobo¹

Introdução

Ao longo da minha carreira docente, debati-me várias vezes com a pergunta seguinte: será possível haver uma prática letiva de qualidade quando se está sujeito a exame nacional?

Sustentada na tese de doutoramento intitulada *Avaliação Formativa, Avaliação Sumativa e Exames na Prática de Três Professoras de Português de 12.º Ano* encontro uma resposta afirmativa. Fundamenta-a o facto de Beatriz, uma das participantes do estudo, ter vestido, literalmente, a camisola do exame e, à luz do perfil do professor, definido na tese, também *ter vestido*, em aula, *a camisola da avaliação para as aprendizagens*. Esta complementaridade foi conseguida, na medida em que a docente, imbuída das características do professor reflexivo, conjugou sabiamente a avaliação formativa com a sumativa, como revelam os *avaliogramas* desenhados. Demonstrou que ambas podem coabitar em aula, mesmo com alunos que serão sujeitos a exames nacionais. Porém, fica algum incómodo pelo facto de classificações de exame terem sido inferiores ao esperado.

Este texto começa por dar conta dos resultados da investigação, para depois referir algumas práticas consideradas significativas, forma de fundamentação da resposta à questão inicial.

A investigação

A tese de doutoramento *Avaliação Formativa, Avaliação Sumativa e Exames na Prática de Três Professoras de Português de 12.º Ano*, defendida em 2010, no Instituto de Educação da Universidade de Lisboa, resulta duma investigação de índole qualitativa, de inspiração etnográfica,

¹ Agrupamento de Escolas D. Maria II, Sintra, Unidade de Investigação do Instituto de Educação, Universidade de Lisboa.
Autora da tese de doutoramento aqui referida.

cuja unidade de análise é o professor, suas ações e pensamentos. Não se pode, por isso, generalizar exclusivamente com estes dados.

O enquadramento teórico baseia-se na discussão do perfil do professor (relativamente à prática letiva e avaliativa), de autores consagrados, quer da área dos saberes docentes (como Clark e Hollingsworth, 2002, Hargreaves, 1994, 2000, Huberman, 2000, Joice e Calhoun, 2009, Korthagen, 2005, Schön, 1983, Shulman, 1987, Zeichner, 1993), quer do domínio da avaliação dos alunos (como Black e William, 1998, Crooks, 1988, Fernandes, 2005, Gipps, 2004, Goodlad, 2004, Harlen, 2006, Natriello, 1987, Shepard, 2001, Stiggins e Conklin, 1992, Taras, 2005).

As professoras e respetivas turmas da Grande Lisboa foram selecionadas pelos órgãos de gestão, em função dos critérios que entenderam. A turma 12A foi indicada por ser, sem dúvida, a melhor turma da escola A, pública, e pelo facto de a professora Amélia não se importar de participar no estudo. Beatriz foi considerada uma inevitabilidade, já que era a única professora de 12º ano da escola B, privada; a turma 12B surgiu por acaso, pois não havia diferença significativa entre elas, tanto ao nível dos resultados como do comportamento. Carolina, da escola C, pública, encontrava-se numa fase profissional de desilusão, pelo que viu com agrado o aparecimento de um olhar externo que pudesse ajudá-la a compreender esse sentimento; das suas turmas, a 12C *ainda era das melhorzinhas*.

A recolha de dados foi equilibrada nas três escolas e decorreu, essencialmente, ao longo de 2006/2007. Baseou-se na observação de 86 aulas. De todas, foram elaboradas tanto sínteses descritivas, que evidenciaram os parâmetros em observação, como avaliogramas (assim designados pela semelhança com os eletrocardiogramas), que permitiram visualizar de outro modo o *sentir da aula*. Baseou-se também em mais de 13 horas de entrevistas integralmente transcritas e interpretadas, feitas às professoras, a um elemento da direção da escola e a alunos. Apenas as dos alunos foram entrevistas coletivas. Baseou-se ainda na análise de conteúdo do diário de campo e de artefactos escritos, estes num total de 154 documentos (documentos estruturantes das escolas, instrumentos de

avaliação, resolução das provas de exame de Português B dos alunos entrevistados, etc.).

A informação foi tratada de modo a perceber-se o tipo de relação existente entre a avaliação formativa e a sumativa, o impacto dos exames na prática letiva, a forma de gestão quer das pressões dos exames quer da aplicação das duas modalidades de avaliação. Assim, veria se é possível haver uma prática letiva de qualidade quando se está sujeito a exame nacional.

Como afirmado na introdução, a minha resposta é afirmativa, pois identifiquei pelo menos um caso que revelou que a avaliação externa não comprometeu a prática letiva e avaliativa de qualidade. E se existe um caso é porque é possível! No entanto, as classificações de exame ficaram aquém dos desempenhos revelados pelos alunos ao longo do ano. Assento, pois, a minha fundamentação nas evidências, tanto nos resultados dos exames como nas práticas observadas sobretudo em Beatriz e em todo o contexto da turma 12B.

As evidências

Num ano em que a média nacional das classificações de exame (CE) de 1ª fase se situou nos 10,8 valores, podemos dizer que os resultados das três turmas, revelados na Tabela 1, foram bons. O incómodo atrás aludido surge por duas razões: a) pela inesperada descida da média da classificação final da disciplina no 12º ano (CF) da turma 12B (diferença de 4,41 valores para CE) e b) pela inversão das médias de exame das turmas 12A e 12B, quando as observações identificaram estes últimos alunos com um desempenho bastante superior ao dos primeiros.

Tabela 1. Valores médios de classificação de frequência (CF), classificação interna final (CIF) e classificação de exame (CE) comparados nas três turmas

	CF	CIF	CE	CE-CF
A	15,88	14,88	14,11	-1,77
B	17,88	16,70	13,47	-4,41
C	12,84	13,07	11,38	-1,30

Importa também referir duas das conclusões do estudo que vêm a propósito. Por um lado, identificaram-se alguns efeitos de *washback*, i.e., pressões sentidas pela existência de exames, como a alteração consciente e inconsciente das classificações internas, que aumentam a média de classificação interna final (CIF) – procedimento verbalizado por Amélia e subentendido em Carolina. Por outro lado, os exames não se afiguraram determinantes na prática letiva das professoras, que parece estar mais dependente dos seus estilos.

Relativamente à prática letiva, da análise da literatura, identifiquei as características fundamentais que permitem definir dois perfis extremados de professores. Considerei que, para cada parâmetro, um professor pode situar-se num ponto de uma linha contínua que vai de um estilo mais técnico a outro mais reflexivo. Diretamente relacionados com as práticas de ensino, encontram-se os parâmetros: abordagem dos conteúdos (que pode ser mais superficial ou mais profunda), recurso a tarefas (mais simples e rotineiras ou mais complexas e diversificadas), processo de interação oral (mais espontâneo e intuitivo ou mais deliberado e reflexivo), dinâmicas de sala de aula (mais individualizadas ou mais cooperativas) e exploração do currículo (mais restritiva ou mais abrangente). Nas ligadas às práticas avaliativas, temos os parâmetros: finalidades da avaliação (mais sumativas ou mais formativas), tipo de *feedback* (mais classificativo ou mais orientador), participação dos alunos (mais passiva ou mais ativa) e preparação para os exames (mais descontextualizada ou mais contextualizada). A síntese destes nove parâmetros conduz a um último que identifica o estilo de determinado professor como mais técnico ou mais reflexivo. Das evidências recolhidas resultou a distribuição constante da Tabela 2.

Tabela 2. Sistematização do perfil de ensino e avaliação de Amélia, Beatriz e Carolina

Práticas de ensino	Abordagem dos conteúdos	+ superficial				Carolina	Amélia				Beatriz	+ profunda
	Recurso a tarefas	+ simples e rotineiras		Amélia		Carolina					Beatriz	+ complexas e diversificadas
	Processo de interação oral	+ espontâneo e intuitivo		Amélia		Carolina					Beatriz	+ deliberado e reflexivo
	Dinâmicas de sala de aula	+ individualizadas		Amélia			Carolina	Beatriz				+ cooperativas
	Exploração do currículo	+ restritiva		Amélia			Carolina				Beatriz	+ abrangente
Práticas avaliativas	Finalidades da avaliação	+ sumativas		Amélia		Carolina				Beatriz		+ formativas
	Tipo de <i>feedback</i>	+ classificativo				Amélia		Carolina	Beatriz			+ orientador
	Participação dos alunos	+ passiva			Carolina	Amélia					Beatriz	+ ativa
	Preparação para os exames	+ descontextualizada				Amélia	Carolina	Beatriz				+ contextualizada
Estilo de ensino e avaliação	+ técnico			Amélia	Carolina					Beatriz	+ reflexivo	

Mesmo sem entrarmos em pormenores, consegue perceber-se que Beatriz (situada à direita) é, das três intervenientes, a que melhor integra o perfil reflexivo, o defensável no paradigma atual. Por esta razão, fundamento a minha resposta inicial na sua prática, que apresento de imediato, parcial e sucintamente.

Segundo os alunos, Beatriz é expressiva, enérgica, direta, dura e exigente. As suas aulas têm um ritmo próprio e acelerado. Eles apreciam-nas por serem originais e criativas. As *atividades lúdicas*, como lhes chama o Bruno (EABruno,1), não tornam as aulas maçadoras e ajudam-nos a compreender a poesia, modo literário de que não gostavam no início, mas que já aprenderam a gostar. Reconhecem a preocupação constante da professora em criar surpresa, em revelar curiosidades relacionadas com os conteúdos, o que permite dizerem que *chega até a dar vontade de ficar mais noventa minutos*; sentem-se motivados.

Cada aula é introduzida por uma conversa informal, em princípio externa ao objetivo do Português. Depois, faz, de forma mais ou menos subtil, uma revisão das últimas matérias, com o intuito de obter “*feedback* do que lhes ficou mesmo sem eles terem estudado” (EPB3,10). Este diagnóstico das aprendizagens aparece integrado num discurso fluido, como parte

natural do diálogo inicialmente estabelecido e que vai *resvalando* para os conteúdos temáticos, que explora durante a aula. Trata-se, de resto, de um procedimento habitual na articulação entre as diferentes partes destas aulas (OAB5). Beatriz fá-lo conscientemente.

A professora possui um conhecimento profundo da matéria que leciona, muito além da informação dos manuais escolares. Os dados novos surgem em curtas exposições, entrecortadas pelo diálogo, por declamações, por leituras (normalmente feitas pela professora), por relatos de histórias curiosas desconhecidas da turma e que prendem a sua atenção. Relaciona-os com a atualidade.

Beatriz tem uma forma *adulta* de lidar com os alunos. Por um lado, a exigência que lhes pede tem-na para consigo própria, quer pelo modo como aprofunda as temáticas quer pelo vocabulário específico que usa na exposição e na relação dialógica que estabelece quer ainda pelo que espera do desempenho discente. As correções que faz e as fundamentações que pede sistematicamente são disso testemunho. Por outro lado, aceitar que o conhecimento se constrói com base em múltiplas interpretações dá-lhe uma humildade, que se sente verdadeira. Esta humildade materializa-se no reconhecimento da sua incapacidade em responder a dúvidas sobre a Terminologia Linguística para os Ensinos Básico e Secundário (TLEBS) ou na hipótese de que a falha de comunicação possa estar do seu lado (“Não foi exatamente isso que eu disse... Eu também posso não ter sido muito clara” - OAB9) ou ainda na genuína valorização da opinião dos alunos, aceitando e estimulando a sua perspetiva na interpretação de textos, o que foi muito frequentemente observado. Esta abertura de espírito já *desencantou* alguns alunos por acreditarem que o professor deve saber tudo, todavia, diz não saber agir de outro modo.

A metodologia interativa a que recorre constitui outra das suas mais-valias, pela forma hábil como a implementa: faz perguntas cuja formulação exige respostas abertas; atribui tempo necessário à elaboração do raciocínio, sem quebrar o ritmo da aula; é perseverante nos pedidos de fundamentação e de exemplificação, estimulando os pensamentos de nível

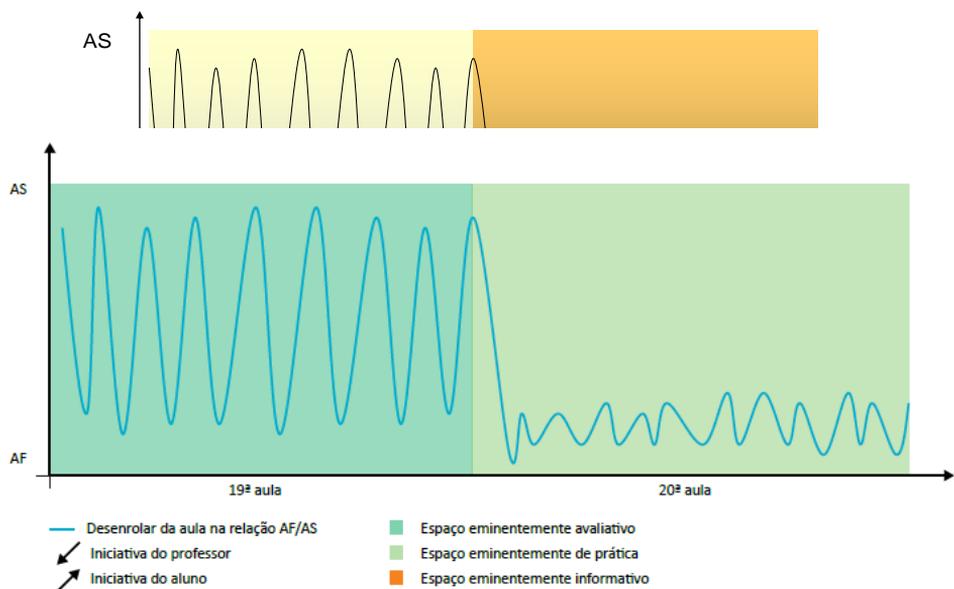
superior; incute autoconfiança nos adolescentes pela valorização das suas opiniões; fornece *feedback* constantemente.

O currículo foi desenvolvido em praticamente toda a sua extensão (contrariamente ao observado nas outras professoras), incluindo apresentações orais e contratos de leitura. As apresentações, bipartidas, constituíram verdadeiros espetáculos: primeiro apresentaram-nos, depois fundamentaram todos os pormenores.

Do ponto de vista das práticas avaliativas, Beatriz revela valorizar muito a avaliação *para* as aprendizagens. Preocupa-se especialmente com a formação dos alunos enquanto cidadãos, por isso quer que compreendam os fenómenos e que fundamentem as opiniões/observações, dá-lhes *feedback* orientador, diversifica as tarefas, insistindo nas metacognitivas. Contudo, preocupa-se igualmente com a avaliação sumativa: apela à realização voluntária de textos de diversas tipologias, que coloca na *internet*; disponibiliza-se para os corrigir e comentar, nomeadamente nas aulas semanais de esclarecimento de dúvidas; perante a indecisão da tutela na manutenção da TLEBS, acaba por trabalhá-la nas últimas aulas, pois pode constar do exame...

Mesmo ao nível da avaliação interna, a sumativa é bastante ponderada. Da articulação que a docente fez entre ambas as modalidades de avaliação, destaco uma revelada na 19^a aula, constante do avaliograma da Figura 1. As duas modalidades complementaram-se.

Eis a situação: Beatriz tinha solicitado a cada grupo de alunos uma nova capa para a obra *Felizmente Há Luar!* Já as tinha recebido, analisado e anotado. Porém, sentia-se incapaz de atribuir classificações por falta de informação que sustentasse os resultados. Em aula, os alunos justificaram cada símbolo usado, a professora foi fazendo perguntas e eu assisti à alternância, num largo espetro, entre a profundidade típica da avaliação formativa e a intencionalidade sumativa da avaliação.



Beatriz sabe que os exames constituem um momento particularmente importante na vida dos adolescentes e famílias. Solidariza-se com eles. De tal modo que, não só os estimula ao máximo como, no dia do exame, enverga uma t-shirt com Fernando Pessoa, à frente, e *Exame Nacional de Português*, atrás. Beatriz veste a camisola do exame! E pelo seu empenho, não tenho dúvidas que também *veste a camisola da prática letiva!*

Porém, *nem tudo foram rosas*. Beatriz viveu incrédula e angustiada pelos resultados de exame que não corresponderam ao esforço de ambos os lados (professora e alunos). Logo em julho, quando as classificações da primeira chamada foram publicadas, a desilusão pelos resultados de Português instalou-se. Quase todos os alunos pediram recurso, que foi total ou parcialmente aceite. Com a publicação dos *rankings* nacionais, a docente reforçou esse sentimento, refletindo: se, por um lado, os alunos são bons e também ela cumpre o seu dever, por outro, os resultados que obtiveram nos exames não foram satisfatórios. Porquê? Na sua escola, Português foi a disciplina com classificações mais baixas! Nos *rankings* só a Português, ficou em 90º lugar! O que é que está mal?

Reflexões finais

As respostas às questões que Beatriz levantou exigiriam uma nova investigação. Todavia, não tenho dúvida de que algo está mal! Reveja-se a Tabela 1. Pelo desempenho revelado ao longo do ano e pelo confronto entre as provas de exame, inquestionavelmente, a turma 12B merecia: ter melhores classificações; ter classificações superiores à 12A.

Ocorreram-me, quatro hipóteses justificativas:

1. Alinhamento entre o desenvolvimento do currículo e o exame – porque obteve melhores resultados a turma que viu o currículo prescrito significativamente reduzido e não beneficiou, em aula, de uma avaliação para as aprendizagens? Terá isso a ver com a natureza do exame aplicado? Será que desenvolver uma diversidade de competências e conhecimentos previstos, mas não avaliados na prova nacional, não melhora a qualidade das respostas? Ou a dispersão dos alunos pelo desenvolvimento das diferentes competências funciona como distrator da preparação para o exame? De acordo com Popham (2008) estes resultados não serão estranhos. O autor contesta a asserção de Black e Wiliam (1998), não porque uma melhor aprendizagem não melhore os resultados, mas por considerar os exames de controlo das aprendizagens insensíveis à lecionação: eles não detetam a diferença entre uma formação adequada e uma inadequada. Popham (2008) defende que as provas testem um número modesto de objetivos, que sejam claramente descritas as competências a serem testadas e que se faça um relatório por aluno para cada um desses objetivos. Só assim os exames poderão ser sensíveis à lecionação. Mas poderá esta proposta funcionar? Como fazer, então, alinhar a avaliação externa com a visão reformista do currículo?
2. Classificação dos exames – para além dos critérios distribuídos aos professores-classificadores, que outros critérios, subjetivos, entram em jogo? A questão não é simples, porquanto, ultimamente, tem-se assistido a sucessivas alterações metodológicas. Porém, a Português, a escala de classificação raramente é usada em toda a amplitude, situação

que pode beneficiar os alunos muito fracos e prejudicar os muito bons. Existe, por isso, a tendência para se descontar algo quando tudo se apresenta bem (ou para contar algo quando a resposta é muito fraca), i.e., há a tendência para a centralidade das classificações (Noizet e Caverni, 1978), situação a que todos os alunos estarão sujeitos. Mas pensemos: será que os alunos de uma escola muito boa, cujas provas pertencem todas ao mesmo lote (muito bom, portanto), não são prejudicados por esse mesmo facto, quando cotejados com outros que, sendo também muito bons, têm as suas provas misturadas com outras boas e satisfatórias? Poderão estes efeitos de ancoragem estar a jogar o seu papel no caso aqui investigado?

3. Autoconfiança e ansiedade dos alunos – surgem-me algumas dúvidas respeitantes à influência das diferentes práticas pedagógicas: que fatores mais contribuem para resultados mais eficazes? Será que um professor bastante rigoroso e exigente provoca insegurança nos alunos? Ou será que o facto de os alunos perceberem que há sempre algo a melhorar (mesmo com uma sólida formação) lhes coarta essa confiança? Será que a ansiedade dos alunos aumenta exageradamente perante expectativas e ambições elevadas (ARG, 2002)?
4. Formação global prévia ao 12º ano – levanta-se a questão: será que as aprendizagens anteriores ao ano de exame, adquiridas ao longo do ciclo, exercem um peso bastante superior ao imaginado e, conseqüentemente, as aprendizagens de 12º ano não são tão significativas? Ou nesta altura já só os conteúdos são significativos? Ou terá a 12A capacidades e competências que não foram demonstradas em aula de Português apenas porque não foram explicitamente solicitadas?

Tendo começado com uma pergunta, acabo com muitas outras. Há, assim, espaço para novas investigações!

Referências

ARG, (2002a). *Testing, Motivation and Learning*. Cambridge: University of Cambridge Faculty of Education.

- Black, P., e Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education*, 5(1), 7-74.
- Clarke, D. e Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, 18, 947-967.
- Crooks, T. (1988). The Impact of Classroom Evaluation Practices on Students. *Review of Educational Research*, 58(4), 438-481.
- Fernandes, D. (2005). *Avaliação das Aprendizagens: desafios as teorias, práticas e políticas*. Cacém: Texto Editores.
- Gipps, C. (2004). *Beyond Testing: Towards a Theory of Educational Assessment*. Oxon: Routledge Falmer.
- Goodlad, J. (2004). *A Place Called School*. Nova Iorque: McGraw-Hill. Edição especial do vigésimo aniversário.
- Hargreaves, A. (2000). Four Ages of Professionalism and Professional Learning. *Teachers and Teaching: History and Practice*, 6(2), 151-182.
- Hargreaves, D. (1994). The New Professionalism: the synthesis of professional and institutional development. *Teaching & Teacher Education*, 10(4), 423-438.
- Harlen, W. (2006). On the Relationship Between Assessment for Formative and Summative Purposes. In Gardner (Ed.) *Assessment and Learning*. Londres: Sage Publ., pp. 103-117.
- Huberman, M. (2000). O Ciclo de Vida Profissional dos Professores. In Nóvoa (org.) *Vidas de Professores*. Porto: Porto Editora, 2ª ed., pp. 31-62.
- Joyce, B. e Calhoun, E. (2009). Three Sides of Teaching: Styles, Models, and Diversity. In Saha e Dworkin, *International Handbook of Research on Teachers and Teaching*. Nova Iorque: Springer, pp. 645-652.
- Korthagen, F. A. (2005). Practice, Theory and Person in Lifelong Professional Learning. In Beijaard, Meijer, Morine-Dershimer e Tillema (Eds.). *Teacher Professional Development in Changing Conditions*. Holanda: Springer, pp. 79-94.
- Natriello, G. (1987). The Impact of Evaluation Processes on Students. *Educational Psychologist*, 22(2), 155-175.
- Noizet, G. e Caverni, J. (1978). *Psychologie de l'évaluation scolaire*. Paris: PUF.
- Popham, J. (2008). *Transformative Assessment*. Virginia, EUA: Association for Supervision and Curriculum Development.
- Schön, D. (1983). *The Reflective Practitioner*. Nova Iorque: Basic Books.
- Shepard, L. A. (2001). The Role of Classroom Assessment in Teaching and Learning. In Richardson (Ed.) *The Handbook of Research on Teaching* (4ª ed., pp. 1066-1101). Washington, DC: American Educational Research Association.

- Shulman, L. (1987). Knowledge and Teaching: Foundations of the New Reform. *Harvard Educational Review*, 57(1), 1-22.
- Stiggins, R. e Conklin, N. (1992). *In Teacher's Hands*. Nova Iorque: State University of New York Press.
- Taras, M. (2005). Assessment – Summative and Formative – Some Theoretical Reflections. *British Journal of Educational Studies*, 53(4), 466-478.
- Zeichner, K. (1993). *A Formação Reflexiva de Professores: Ideias e Práticas*. Lisboa: Educa.

Descodificação das fontes

EABruno,1 – entrevista ao aluno Bruno, pág. 1

EPB3,10 – terceira entrevista à professora Beatriz, pág.10

OAB5 – observação da 5ª aula de Beatriz

OAB9 – observação da 9ª aula de Beatriz

Introdução

A avaliação externa dos alunos, concretizada em testes intermédios e, em particular, em exames nacionais, é um instrumento regulador das suas aprendizagens. É, também, um meio de seleção para o ingresso no ensino superior.

Por isso tornou-se, hoje, um tema central de reflexão e de preocupação nas escolas. Tal facto tem necessariamente repercussões na gestão do currículo e na prática letiva do professor, ora com consequências positivas ora originando constrangimentos.

Aspetos positivos da avaliação externa dos alunos na prática docente

Destacam-se como aspetos positivos da avaliação externa dos alunos na prática docente os seguintes:

- maior responsabilização dos professores na gestão do currículo, levando a uma preparação mais cuidada do processo de ensino e aprendizagem e maior reflexão sobre o mesmo;
- aumento do trabalho colaborativo entre professores;
- cumprimento dos programas, em particular no que respeita à lecionação dos conteúdos;
- maior assiduidade dos professores.

A avaliação externa dos alunos é vista, também, como uma forma de avaliar o trabalho dos professores e das escolas. Por isso nas escolas, e em particular nos grupos disciplinares, há uma grande preocupação pela gestão do currículo das disciplinas dos anos com exames nacionais, o que nem sempre acontece com as dos restantes anos de escolaridade. Por exemplo, no grupo disciplinar de Física e de Química, as disciplinas de Física e de Química do 12.º ano, que antigamente tinham um lugar de

¹ Escola Secundária Frei Heitor Pinto, Covilhã.

destaque no currículo, passaram a ser vistas como disciplinas “menores”, especialmente pelos alunos, pelo facto de já não terem exame nacional.

As disciplinas em que há exames nacionais tornaram-se, portanto, o alvo principal de tomada de decisões sobre a gestão do currículo em reuniões de grupos disciplinares e departamentais. Tal facto tem conduzido a uma promoção do trabalho colaborativo entre professores, em especial no que concerne à preparação de materiais, com relevância positiva para professores menos experientes. Os baixos resultados nos exames nacionais também levaram os professores a refletirem mais sobre o processo de ensino e aprendizagem, designadamente através da análise de relatórios sobre o desempenho dos alunos na avaliação externa, na expectativa de uma melhoria dos resultados.

O cumprimento dos programas também é um aspeto positivo resultante da avaliação externa. Na antiga disciplina de Física e Química dos 10.º e 11.º anos não havia exames nacionais; verificava-se que, em muitos casos, havia conteúdos que não eram lecionados numa das componentes (Física ou Química). Tal situação tinha repercussões negativas no prosseguimento de disciplinas da mesma área do 12.º ano e do ensino superior. Hoje isso não se verifica, todos os conteúdos são abordados. Relativamente às metodologias propostas nos Programas, elas poderão ser implementadas em maior ou menor extensão e qualidade, mas isso também acontecia quando não havia exames nacionais, uma vez que é um aspeto que depende essencialmente do professor.

A extensão dos programas e o seu cumprimento integral levou a uma maior assiduidade por parte dos professores, embora algumas vezes de uma forma negativa: a preocupação com as disciplinas de exame é tão grande que os professores acabam por comparecer nas aulas mesmo doentes.

Aspetos negativos da avaliação externa dos alunos na prática docente

Apontam-se alguns aspetos negativos decorrentes da existência de avaliação externa:

- tendência para uma maior ênfase no desenvolvimento de competências e conhecimentos objeto de avaliação em exames, subvalorizando-se outros;
- risco de um aprofundamento desadequado dos conteúdos, promovendo a desistência de alunos com mais dificuldades e, portanto, um ensino para elites;
- maior controlo/pressão sobre o professor para o cumprimento de calendarizações (pelos pares, alunos, encarregados de educação, direção) com possíveis consequências no ritmo e qualidade das aprendizagens;
- risco de uma padronização de metodologias e instrumentos de avaliação interna;
- efeitos da cultura da escola (pressão dos *rankings*, da avaliação externa das escolas, dos contratos de autonomia): ensinar para o sucesso na avaliação externa? Ou ensinar para o sucesso na avaliação interna?

Os professores, pressionados pela melhoria dos resultados nos exames nacionais e pela falta de tempo em abordar com qualidade os Programas, tendem a selecionar os conhecimentos e, em particular, as competências que são objeto de avaliação em exame. Tal atitude subvaloriza o desenvolvimento e respetiva avaliação de aspetos dos Programas que contemplam uma formação mais geral do aluno no âmbito disciplinar (por exemplo, a pesquisa e análise documental e a comunicação oral de ideias com utilização de linguagem própria da disciplina).

Também alguns professores, receosos do grau de exigência das provas de exame, tendem a aprofundar em demasia certos conteúdos. Tal atitude torna a disciplina inacessível para os alunos que têm mais dificuldades, podendo promover a sua desistência.

Há também um grande controlo sobre o professor relativamente ao cumprimento da calendarização anual feita no início do ano letivo. Ele vem não só da direção da escola mas, particularmente, dos pares (o cumprimento da calendarização é visto como sinónimo de bom profissionalismo), como dos próprios alunos e encarregados de educação através de manifestações de preocupação sobre a possibilidade do

professor não lecionar atempadamente os conteúdos. A gestão do currículo passou, portanto, a extravasar o domínio do professor, estendendo-se a outros agentes educativos. Ora esta pressão, para além de tornar pouco saudáveis as relações institucionais, tende a promover uma uniformização no tratamento dos alunos, pois impõe o mesmo ritmo de aprendizagem a todas as turmas independentemente das suas características. E são os alunos com dificuldades que mais sofrem com esta tentativa de uniformização.

Há igualmente um risco de padronização das metodologias e dos instrumentos de avaliação interna usados, os quais são escolhidos pelo grupo disciplinar tendo em conta as características do exame nacional. Por um lado, a utilização de metodologias semelhantes e, em particular, de instrumentos de avaliação (por exemplo, usar os mesmos testes em todas as turmas do mesmo ano) leva ao tratamento igual de alunos que podem ter características muito diferentes, com eventual prejuízo da qualidade das suas aprendizagens. Por outro lado, as decisões coletivas provenientes de trabalho colaborativo ou de estruturas superiores da escola sobrepõem-se à decisão individual do professor, podendo colocá-lo numa posição de mero técnico que cumpre orientações. O professor, enquanto construtor de currículo, pode ser posto em causa, o que leva a um empobrecimento na diversidade e na inovação.

Também a cultura de escola influencia as decisões dos professores na gestão do currículo. Há escolas que, pela quantidade e qualidade dos seus alunos, podem ser mais exigentes e trabalham com um objetivo: obter uma boa posição nos *rankings*. Outras, pelo contrário, instaladas em zonas desfavorecidas, não têm esta pretensão como objetivo principal, mas sim o sucesso educativo (e não só escolar!), investindo nos seus resultados a nível interno. E o professor tem de adotar um estilo que esteja de acordo com a cultura de escola, o que tem implicações na sua prática letiva.

Outros indicadores externos, como a avaliação externa das escolas e os contratos de autonomia, são fatores de pressão nas decisões dos professores: por um lado há que melhorar o sucesso na avaliação interna, por outro não deve haver grandes discrepâncias entre a avaliação interna e

externa. E a expectativa sobre a conciliação destes resultados recai sobre o professor, aumentando a sua responsabilização no processo de avaliação.

Conclusões

Como conciliar os aspetos positivos e negativos da avaliação externa dos alunos e a gestão do currículo pelo professor?

É incompatível a avaliação externa e a diversidade e inovação na prática letiva do professor?

Nota-se uma pretensão em avaliar os professores pelos resultados que os seus alunos obtêm nas provas de exames nacionais. Esta pretensão, sendo errada, é mais um fator de pressão sobre o professor. Verifica-se que um mesmo professor ora tem bons resultados ora menos bons. Eles dependem de muitas variáveis mas, essencialmente, dos alunos e do tipo de prova. Nem o desempenho de um aluno nem o de um professor pode ser avaliado apenas pelo resultado de um exame.

E nem todo o tipo de conhecimento ou competência de um aluno pode ser avaliado num exame. A atestar isto recorda-se que alguns alunos já foram medalhados em olimpíadas internacionais mas não obtiveram a classificação máxima em exames nacionais. Não é, por isso, razoável que a prática letiva do professor seja liderada por objetivos relacionados só com a avaliação externa, pois é preciso promover um desenvolvimento integral do aluno e esse também deve ser avaliado. Não se podem reduzir os objetivos da avaliação interna aos objetivos da avaliação externa, pois eles são diferentes e, por isso mesmo, a legislação confere-lhes pesos diferentes.

Também não se pode reduzir o papel do professor ao de um mero técnico que cumpre acriticamente orientações superiores ou se limita a usar recursos disponíveis. A padronização do ensino, com o objetivo de uma pretensa igualdade que conduz à melhoria de resultados em exames nacionais, pode pôr em causa a qualidade das aprendizagens. Por um lado trata de igual forma o que é diferente – os alunos – não promovendo metodologias de diferenciação. Por outro, também não promove o

desenvolvimento profissional do professor e põe em causa a sua identidade: é suposto este construir e reconstruir, através da reflexão, ambientes de aprendizagem diversificados e adequados às características dos seus alunos.

A avaliação externa não é incompatível com um professor construtor de currículo. É preciso que haja uma diminuição da tensão entre as decisões coletivas e a decisão individual do professor. Deve haver mais liberdade e confiança no professor relativamente às suas opções didáticas, pois é a este que cabe a tomada de decisão em sala de aula, baseada nas necessidades individuais dos alunos, uma vez que a natureza da aprendizagem requer flexibilidade e capacidade de resposta. É a reflexão constante sobre o seu trabalho e a renovação das suas estratégias que é promotora de ambientes de aprendizagem adequados às necessidades dos alunos. E são esses ambientes que podem promover bons resultados, incluindo os da avaliação externa.

EXAMES NACIONAIS DE BIOLOGIA E GEOLOGIA E AVALIAÇÃO FORMATIVA: PARA UMA RECONCILIAÇÃO

Paula Serra¹ & Cecília Galvão²

A articulação entre avaliação formativa e avaliação sumativa, nomeadamente na modalidade de exames nacionais, tem sido amplamente discutida, na literatura, por vários autores (Wiliam e Black, 1996; Harlen, 2006; Wiliam, 2000, Santos, 2013), sendo frequente considerar-se que estas duas modalidades de avaliação são de difícil compatibilização. Esta dificuldade deve-se, em parte, ao facto de estes dois tipos de avaliação terem funções diferentes: a avaliação formativa tem como propósito promover aprendizagens; a avaliação sumativa tem como função estabelecer uma sinopse das aprendizagens alcançadas. Ou, por outras palavras, a avaliação formativa tem uma natureza *prospetiva*, enquanto a avaliação sumativa tem uma natureza *retrospectiva* (Wiliam, 2000).

Para além da dificuldade de articulação entre avaliação sumativa e formativa, muitos autores defendem, também, que a avaliação sumativa externa pode conduzir ao empobrecimento do ensino, nomeadamente no que respeita à avaliação formativa, como se a primeira pudesse aniquilar a segunda. “O aumento da importância de ‘ensinar para o teste’ trouxe um empobrecimento curricular e o negligenciar de técnicas pedagógicas associadas à avaliação formativa” (Gardner, 2006, p.197). Ou seja, para além de se aceitar que estas duas modalidades de avaliação são incompatíveis, parece existir a ideia de que, em certo sentido, nem é desejável que tal articulação ocorra.

Se, por um lado, a investigação educacional tende a valorizar a avaliação formativa, desenvolvida em sala de aula, como uma estratégia eficaz na promoção das aprendizagens; por outro lado, a política educativa tende a

¹ Escola Secundária de Vergílio Ferreira e Unidade de Investigação e Desenvolvimento em Educação e Formação, Instituto de Educação, Universidade de Lisboa.

² Instituto de Educação, Universidade de Lisboa.

valorizar a avaliação sumativa externa e a publicação do *ranking* de escolas, como forma de regulação, monitorização e, até mesmo, como fomentadora da qualidade da educação. Como acontece com quase todas as posições dicotómicas tomadas na área da educação, aceitar que uma das modalidades de avaliação é “melhor” ou mais desejável do que a outra, ou que as duas são incompatíveis, pode ter consequências nefastas. Como diz Nóvoa (2010, p. 40), “Para aceder à inteligência das coisas não podemos ceder ao facilitismo do pensamento binário.”

Neste texto, argumenta-se que as tarefas de avaliação incluídas nos exames nacionais de Biologia e Geologia (BG) podem ser usadas pelos professores, em sala de aula, para desenvolver avaliação formativa, funcionando, assim, como instrumentos pedagógicos charneira, capazes de articular a avaliação formativa e sumativa. Ou seja, servindo, em simultâneo, para sustentar a progressão das aprendizagens dos alunos e para a obtenção de melhores resultados escolares. O argumento estrutura-se, ao longo do texto, em três componentes: 1) Desmistificações da ideia de que os professores não devem ensinar para os exames; 2) Identificação de um problema – os maus resultados obtidos pelos alunos portugueses nos exames nacionais de BG –, estabelecimento de hipóteses explicativas e propostas de atuação e 3) Descrição de uma intervenção educativa, implementada com vinte professores do ensino secundário, que parte de questões de exame para o desenvolvimento de estratégias de avaliação formativa focadas no desenvolvimento do raciocínio científico e na construção de explicações científicas.

1. Mito: Os professores não devem ensinar para os exames

A ideia de que os professores não devem ensinar para os exames tem sido, na área da educação em ciência, uma fonte de tensão entre investigadores e professores. Por um lado, os investigadores, centrados na literacia científica, tendem a defender que ensinar para os exames promove aprendizagens não significativas, rapidamente esquecidas após a realização da prova. Isto ocorrerá porque o formato tradicional dos exames favorece a memorização e a compreensão simples de factos e conceitos,

avaliadas através de itens de resposta fechada, que permitem objetivar a comparação e seriação do desempenho dos alunos. Assim, atividades de *inquiry*, argumentação e discussão, por exemplo, tão importantes para o desenvolvimento de competências científicas, seriam preteridas em sala de aula, empobrecendo o ensino. Por seu turno, os professores, centrados no sucesso dos alunos nos exames, tendem a defender que ensinar para os exames promove as aprendizagens realmente valorizadas pelo Ministério da Educação e Ciência e que o que importa é o acesso dos alunos ao ensino superior.

Estes dois objetivos – literacia científica e sucesso nos exames –, ambos legítimos e desejáveis, não são incompatíveis. É possível ensinar para os exames, sem comprometer os princípios orientadores da educação em ciência, desde que estes estejam articulados.

Vários autores (James, 2006; Pellegrino, 2009) defendem que a articulação entre currículo, ensino e avaliação pode ser conseguida se todas estas dimensões da educação derivarem da mesma conceção de cognição e aprendizagem. Assim, uma teoria de aprendizagem, cientificamente credível, constitui o elemento chave que permite que as três dimensões “sejam dirigidas para os mesmos fins e se potenciem mutuamente, em vez de trabalharem com objetivos opostos” (Pellegrino, 2009, p. 5).

Em Portugal, é possível observar como diferentes conceções de aprendizagem das ciências têm afetado, ao longo do tempo, o currículo, o ensino e os exames de Biologia e de Geologia. Não sendo visões dicotómicas ou exclusivas, podemos identificar uma visão mais cognitivista e uma visão mais construtivista da aprendizagem.

Numa visão mais cognitivista da aprendizagem, são enfatizados a mente e o processamento de informação. O currículo e a avaliação organizam-se em torno de conhecimentos e processos cognitivos, expressos através de objetivos educacionais. No ensino são privilegiadas metodologias baseadas na transmissão, no fluxo e processamento de informação, nas ideias prévias dos alunos. Nos exames, as questões surgem

descontextualizadas, já que o que importa é a complexidade do conhecimento e dos processos cognitivos a que apela cada item. Uma questão deste tipo poderá, por exemplo, solicitar ao aluno para observar a figura de um estágio de desenvolvimento embrionário de uma ave e indicar o nome e função de determinada estrutura, presente nessa figura. A figura é fornecida sem qualquer enquadramento e, para responder corretamente, o aluno terá, necessariamente, que ter tido contacto prévio com figuras semelhantes à fornecida na prova. Este era o modelo de itens mais visível nos exames de Biologia e de Geologia até 2006.

Numa visão mais construtivista da aprendizagem, ou de cognição situada, para além da mente, importa o sujeito e o contexto. O currículo organiza-se em torno de situações de aprendizagem (Joannaert et al., 2006), dando-se ênfase à perceção/ação, ou seja à “leitura” que o aluno é capaz de fazer da situação e à capacidade de agir sobre ela, através da mobilização das competências apropriadas. No ensino são privilegiadas metodologias centradas em projetos, trabalho prático de natureza investigativa, estudos de caso, discussões... Ou seja, situações de aprendizagem que se aproximem das práticas científicas ou que coloquem os alunos perante problemas do quotidiano que exijam uma interpretação científica. Nos exames, as questões surgem enquadradas por *cenários científicos*, que podem ser casos reais de investigação ou situações concretas da natureza, com os quais os alunos nunca tiveram contacto. Uma tarefa de avaliação deste tipo poderá, por exemplo, partir do resumo de um artigo científico, fornecendo-se ao aluno o objetivo da investigação, procedimento, técnicas de obtenção de dados e resultados experimentais. Deste cenário científico decorre um conjunto de questões, entre as quais se pode incluir, por exemplo, a solicitação de uma explicação dos resultados obtidos na experiência. Os exames de BG são, desde 2006, deste último tipo, sendo construídos em torno de quatro cenários científicos.

Concluindo, i) nalguns aspetos, as tarefas de avaliação presentes no atual modelo de exames de BG, estão mais próximas de situações de investigação científica autêntica, nas quais se pretende envolver os alunos, do que muito do trabalho prático implementado em sala de aula,

considerando as limitações materiais, técnicas e humanas de uma escola; ii) genericamente, as tarefas de avaliação dos exames de BG estão articuladas com os princípios teóricos defendidos para o ensino das ciências, nomeadamente com abordagens construtivistas e baseadas no *inquiry*. Assim, as tarefas de avaliação dos exames nacionais de BG podem constituir-se, a par de outras atividades de sala de aula, como um contexto favorável ao desenvolvimento de competências envolvidas na investigação científica autêntica (*inquiry based learning*) e na literacia científica. Encerram, ainda, a importante vantagem de serem reconhecidas por professores e alunos como relevantes na obtenção de sucesso, o que pode aumentar o seu comprometimento com as tarefas.

2. Problema: Os resultados dos alunos no exame nacional de Biologia e Geologia

Segundo os relatórios anuais publicados pelo Júri Nacional de Exames e pelo Gabinete de Avaliação Educacional, os resultados obtidos pelos alunos no exame nacional de BG têm sido sistematicamente maus, sendo que, nos últimos sete anos, a média nacional de classificações se situa em torno dos nove valores (Sousa et al., 2013).

Pacificada a ideia de que os professores podem usar os exames nacionais como instrumento pedagógico em sala de aula, sem que isso comprometa os princípios orientadores da educação em ciência, importa agora procurar compreender que fatores poderão explicar o fraco desempenho dos alunos no exame de BG. Só assim será possível encontrar vias de atuação que contrariem a tendência verificada nos resultados. Vários poderão ser esses fatores, no entanto dois deles parecem-nos, à partida, particularmente relevantes: i) desadequação das aprendizagens-foco e do grau de exigência cognitiva do instrumento de recolha de dados/exame e ii) desadequação do ensino levado a cabo pelos professores face às competências requeridas pelo exame. Atentemos ao primeiro destes fatores, já que, do segundo, nos ocuparemos na terceira parte deste texto.

Tradicionalmente, o grau de exigência cognitiva dos exames é avaliado tendo em conta critérios relacionados com o grau de complexidade dos

conhecimentos e das capacidades cognitivas mobilizados por um determinado item de avaliação. Para o efeito, podem ser usadas diferentes taxonomias de objetivos educacionais, sendo que uma das mais divulgadas é a Taxonomia de Bloom modificada (Krathwohl, 2002). Num estudo recente, desenvolvido por Ferreira e Morais (2013), foi avaliado o grau de exigência conceptual das questões sobre trabalho prático presentes nos exames e provas intermédias de BG. Usando como critérios de análise, precisamente, a complexidade do conhecimento e das capacidades cognitivas e, ainda, o grau de relação teoria/prática, estas autoras concluem que “A avaliação externa apresenta um baixo nível de exigência conceptual” (p. 20). Parece-nos, contudo, que face ao formato dos atuais exames de BG, alinhado com uma visão de aprendizagem baseada na cognição situada, para além destes critérios ou variáveis mais relacionados com os conhecimentos e capacidades dos alunos, é necessário considerar variáveis contextuais, associadas às tarefas de avaliação (Mislevy, 2008), ou seja relacionadas com os *cenários científicos* apresentados aos alunos.

Mislevy e seus colaboradores, num projeto designado por *PADI — Principled Assessment Designs for Inquiry*, identificam um conjunto de variáveis associadas a diferentes cenários de avaliação de *inquiry*, que designam por *task model variables*. Estas variáveis podem assumir diferentes características que permitem alterar o foco ou dificuldade da tarefa (Baxter e Mislevy, 2005; Riconscente et al., 2005). Tendo em conta o trabalho destes autores e o caso concreto dos exames nacionais de BG, identificam-se algumas variáveis associadas às tarefas de avaliação, que podem condicionar o grau de exigência cognitiva dos exames:

- Grau de familiaridade com a situação/cenário científico (a situação é um exemplo do “padrão” ou uma situação de “exceção” na natureza);
- Linguagem científica (familiaridade do aluno com termos, notações, sinais, símbolos, unidades);

- Grau de delimitação do domínio científico e diversidade dos conceitos envolvidos (de um único domínio científico/inter-relação entre vários domínios científicos);
- Volume de informação;
- Número de distratores. Por exemplo, presença, ou não, de informação não relevante ou contraditória;
- Complexidade do desenho investigativo, em cenários experimentais. Por exemplo, número de variáveis e sua natureza, tipo e diversidade de controlos, familiaridade com as técnicas de recolha de dados e seus fundamentos;
- Complexidade dos resultados experimentais fornecidos. Por exemplo, volume de dados, número de dados confirmativos e infirmativos, facilidade em encontrar padrões/valores significativos/dados anómalos, diversidade e familiaridade com os suportes de dados/representações gráficas (modelos, gráficos, esquemas).

Considerando estas variáveis, uma primeira análise de 83 cenários científicos presentes nos exames de BG, realizados entre 2006 e 2013, leva a crer que estes apresentam, em geral, uma exigência cognitiva considerável.

Tendo em conta a função sumativa dos exames, seria desejável que a avaliação sumativa externa tivesse especial preocupação com a abrangência e representatividade da amostra de itens em cada prova de exame, tendo em conta as aprendizagens estabelecidas no programa. Como diz Pellegrino (2009, p. 24), “Avaliação sumativa em larga escala deve concentrar-se nos aspetos mais críticos e centrais de aprendizagem num determinado domínio, tal como identificados nos padrões curriculares e ser informada pela pesquisa e teoria em cognição”. Por seu turno, a avaliação formativa, em sala de aula, deveria debruçar-se, em profundidade, sobre a qualidade das aprendizagens dos alunos, dando especial atenção ao detalhe, diversidade de situações, situações de exceção, conceitos “fronteira”, ”margens” do programa (de modo a fazer

emergir o que os alunos não sabem e a fragilidade das suas aprendizagens, permitindo ao professor atuar sobre elas).

Ora parece que estas características da avaliação formativa e da avaliação sumativa externa estão invertidas, com os professores a centrarem-se mais nos conceitos e processos cognitivos estruturantes e a avaliação externa a recorrer a uma variedade de estratégias que parecem buscar mais aquilo que o aluno “não sabe”, em vez do que o aluno ”sabe”.

Concluindo, se bem que, em termos genéricos, os exames estejam articulados com os princípios defendidos para a educação em ciência, nomeadamente no que respeita às teorias de aprendizagem mais amplamente aceites, uma análise mais fina revela incongruências que merecem ser reapreciadas, no sentido de se reajustarem as aprendizagens avaliadas e o grau de exigência cognitiva dos exames. Recorrendo, de novo, às palavras de Pellegrino (2009, p. 29 e 30),

“Para que o sistema suporte a aprendizagem deve ter uma qualidade a que chamamos *coerência*. [...] A base conceptual para a avaliação em larga escala deve ser uma versão mais ampla do que faz sentido no nível mais refinado (Mislevy, 1996). Desta forma, os resultados da avaliação externa serão consistentes com a compreensão mais detalhada da aprendizagem resultante do ensino e avaliação realizados em sala de aula.”

3. Ação: uma intervenção educativa para melhorar o raciocínio e a construção de explicações científicas através da avaliação formativa

A desadequação do ensino levado a cabo pelos professores, face às competências requeridas pelo exame, poderá ser um dos fatores que contribui para os maus resultados dos alunos, nos exames de BG. Com as finalidades de alinhar ensino e avaliação e de desenvolver as competências de raciocínio científico e de construção de explicações científicas dos alunos, desenvolveu-se um estudo de intervenção. Partindo de cenários científicos e de questões extraídas de exames de BG, desenharam-se estratégias de avaliação formativa que vinte professores do ensino

secundário, após formação, implementaram em sala de aula, com os seus alunos, ao longo de 5 meses, durante o ano letivo de 2013/2014.

A opção pelo estudo de intervenção justifica-se pelo facto de sermos sensíveis à crítica por vezes apontada à investigação educacional de que esta tem permitido conhecer cada vez melhor a realidade, através de estudos descritivos, mas que pouco tem acrescentado sobre como agir sobre ela, para mudar as práticas dos professores e melhorar as aprendizagens dos alunos, através de estudos de intervenção.

As estratégias de avaliação formativa incidiram sobre o raciocínio científico e a construção de explicações científicas. Estas aprendizagens foram selecionadas porque são transversais e relevantes em qualquer ano de escolaridade e são, segundo os relatórios do GAVE (2008 a 2013), aquelas em que os alunos de BG apresentam maiores dificuldades. Por outro lado, são competências cientificamente relevantes, centrais no empreendimento científico. Como diz Popper (1992, p. 152) “O objetivo da ciência é encontrar explicações satisfatórias do que quer que se nos apresente e nos impressione como estando a precisar de explicação.” As explicações científicas baseiam-se em corpos significativos de conhecimento e de evidências. Por sua vez, “O aspeto central do raciocínio científico consiste na diferenciação e coordenação entre teoria e evidências” (Kuhn, 2010). Assim, quer o raciocínio científico, quer a construção de explicações científicas requerem a articulação entre conhecimento e evidências científicas, podendo ser trabalhadas em simultâneo, em sala de aula.

Muito resumidamente, o estudo desenvolveu-se com base numa metodologia de *design-based research*. Este tipo de metodologia é desenvolvido para testar e refinar desenhos educativos. Pressupõe colocar-se uma primeira versão (protótipo) em sala de aula e otimizá-la, tendo em conta a complexidade do ambiente natural. Considerando as necessidades políticas de implementação de medidas em larga escala, deu-se especial atenção à validade ecológica da intervenção educativa, nomeadamente ao seu potencial de disseminação e amplificação (*scale-up*). Assim, no desenho da intervenção, procurou-se que esta fosse reconhecida como útil

pelos professores (uma das razões pela qual se partiu de questões de exame), não se afastasse demasiado das práticas por eles já desenvolvidas (para que não fosse rejeitada) e, ainda, que não consumisse muito tempo, nem recursos (os principais obstáculos apontados pelos professores na implementação de inovações).

A intervenção educativa compreendeu i) uma formação de professores e ii) uma intervenção pedagógica desenvolvida pelos professores-formandos, durante o normal decorrer das suas aulas. Durante a formação, os professores aprenderam acerca da estrutura de uma explicação científica, como encontrar evidências nos dados de observação ou experimentais, como identificar os principais esquemas de raciocínio científico. Estes conhecimentos científicos (substantivos e epistemológicos) são indispensáveis à interpretação do desempenho dos alunos, quando raciocinam e constroem explicações científicas e ao estabelecimento de um *feedback* de qualidade, realmente impulsionador da progressão das aprendizagens dos alunos, aspetos centrais da avaliação formativa. Durante a formação, os professores construíram explicações científicas e analisaram respostas de alunos, de acordo com um referencial de avaliação, específico para explicações científicas (McNeill e Krajcik, 2008).

A intervenção pedagógica foi desenvolvida em dois ciclos de implementação-reflexão-refinamento de práticas. Ou seja, depois de cada ciclo de intervenção em sala de aula, realizou-se uma sessão de formação conjunta, com a formadora e todos os formandos, para reflexão e apresentação de propostas de refinamento de práticas. Em certo sentido, as atividades desenvolvidas com os alunos mimetizaram as atividades desenvolvidas pelos professores durante as primeiras sessões de formação, organizando-se em cinco momentos chave: i) justificação da importância das explicações científica na atividade académica e científica, ii) clarificação de como se estrutura uma explicação científica, iii) modelação de explicações científicas, pelo professor, iv) construção, de explicações científicas, pelos alunos (individualmente e em grupo), a partir de

questões de exames nacionais, v) questionamento e *feedback* fornecidos pelo professor.

Foram envolvidos nesta intervenção educativa, para além dos 20 professores formandos, cerca de 550 alunos dos 10º e 11º anos. A avaliação do impacto desta intervenção educativa está, ainda, em curso. Contudo, demonstra, desde já, que é possível alinhar a avaliação formativa com a avaliação sumativa externa, em processos de mudança reconhecidos pelos professores envolvidos como muito relevantes.

Conclusão

Se a investigação em desenvolvimento profissional, nos últimos vinte anos, nos mostrou alguma coisa, é que podemos mudar o pensamento dos professores, sem mudar as suas práticas e a única coisa que tem impacto no aproveitamento dos alunos é a prática do professor.

Dylan Wiliam (2007)

A melhoria das práticas de avaliação associadas à aprendizagem das ciências requer um trabalho significativo no que respeita à investigação, formação de professores e disseminação de boas práticas.

Pellegrino (2012) deu o título “*Assessment of Science Learning: Living in Interesting Times*” ao seu comentário final de um número especial da revista *Journal of Research in Science Teaching* sobre avaliação das aprendizagens científicas. Este título enfatiza o facto de vivermos num tempo repleto de incerteza, mas também de oportunidade para a mudança. A avaliação das aprendizagens, sublinha o autor, é um poderoso instrumento para melhorar o ensino das ciências, mas a forma como é considerada nos sistemas educativos pode ter resultados nem sempre desejáveis ou positivos (p. 832).

Procurámos demonstrar neste texto que uma abordagem pragmática da investigação educacional, que tenha em conta as preocupações dos professores e as necessidades políticas, e não perca de vista o que realmente importa (a qualidade das aprendizagens dos alunos), pode

contribuir para se ultrapassarem mitos e problemas e para potenciar as sinergias e o valor educativo da avaliação externa e da avaliação formativa.

Referências

- Baxter, G., & Mislevy, R. J. (2005). *The case for an integrated design framework for assessing science inquiry*. CSE Technical Report 638. Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA. Acedido em http://padi.sri.com/downloads/TR5_IDFramework.pdf
- Ferreira, S. e Morais, A.M. (2013). Exigência conceptual do trabalho prático nos exames nacionais: Uma abordagem metodológica. *Olhar de Professor*, 16 (1), 149-172. Acedido em http://essa.ie.ulisboa.pt/curricula_amorais_texto.htm
- Gardner, J. (2006). Assessment for learning: A compelling conceptualization. In: J. Gardner (Ed.), *Assessment and Learning* (p. 197- 204). London: Sage
- GAVE (Gabinete de Avaliação Educacional, atualmente Instituto de Avaliação Educacional). Relatórios Nacionais de 2008 a 2013. MEC: GAVE. Acedido em <http://www.gave.min-edu.pt/np3/24.html>
- Harlen, W. (2006). On the relationship between assessment for formative and summative purposes (p.103-117). In: J. Gardner (Ed.), *Assessment and Learning*. London: Sage
- James, M. (2006). Assessment, teaching and theories of learning (p. 47- 60). In: J. Gardner (Ed.), *Assessment and Learning*. London: Sage
- Jonnaert, P.; Barrette, J. & Masciotra, D. (2006). *Revisiting the concept of competence as an organizing principle for programs of study: From competence to competent action*. Montréal: Observatoire des réformes en éducation, Université du Québec à Montréal
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, 41(4), 212-218
- Kuhn, D. (2010). What is scientific reasoning and how does it develop? (p.497- 523) In: U. Goswami (Ed.), *Handbook of Childhood Cognitive Development*. West Sussex, UK: Wiley-Blackwell
- McNeill, K. L. & Krajcik, J. (2008). Inquiry and scientific explanations: Helping students use evidence and reasoning(p. 121-134). In J. Luft; R. Bell & J. Gess-Newsome (Eds.), *Science as inquiry in the secondary setting*. Arlington, VA: NSTA.

- Mislevy, R. J. (2008). How cognitive science challenges the educational measurement tradition. *Measurement: Interdisciplinary Research and Perspectives*, 6, 124
- Nóvoa, A. (2010). A terceira margem do rio. *Conferência Que currículo para o Século XXI?*. Lisboa: Assembleia da República/Conselho Nacional de Educação
- Pellegrino, J. W. (2009). *The design of an assessment system for the Race to the Top: A learning sciences perspective on issues of growth and measurement*. Center for K-12 Assessment & Performance Management, Educational Testing Service. Acedido em <http://www.k12center.org/rsc/pdf/PellegrinoPresenterSession1.pdf>
- Pellegrino, J.W. (2012). Assessment of science learning: Living in interesting times. *Journal of Research in Science Teaching*, 49 (6), p. 831-841
- Popper, K. (1992, original publicado em 1956). *O realismo e o objectivo da ciência*. Lisboa: Publicações D. Quixote
- Riconscente, M., Mislevy, R., Hamel, L. (2005). *An introduction to PADI task templates* (PADI Technical Report 3). Menlo Park, CA: SRI International
- Santos, L. (2013). Impacto da avaliação externa nas práticas docentes: Painel III. *Seminário Avaliação das aprendizagens na avaliação da educação*. Lisboa: Conselho Nacional de Educação
- Sousa, H.; Sampaio, M.; Castanheira, M. T.; Pereira, S. e Lourenço, V. (2013). *Análise preliminar dos resultados: Provas finais de ciclo Exames finais nacionais 2013*. Lisboa: GAVE/MEC
- Wiliam, D. & Black, P. (1996). Meanings and Consequences: A Basis for Distinguishing Formative and Summative Functions of Assessment. *British Educational Research Journal*, 22 (5), p. 537-548
- Wiliam, D. (2000, August). *Integrating formative and summative functions of assessment*. Paper presented to Working Group 10 of the International Congress on Mathematics Education, Makuhari, Tokyo
- Wiliam, D. (2007). Content then process: teacher learning communities in the service of formative assessment (p. 183-204). In: D. Reeves (Ed.), *Ahead of the curve: The power of assessment to transform teaching and learning*. Bloomington, IN: Solution Tree

Painel
O QUE FAZER COM OS
RESULTADOS DA AVALIAÇÃO EXTERNA?

Presidente da Mesa – Paulo Sucena

QUALIDADES, DEFEITOS E LIMITAÇÕES DO SISTEMA DE AVALIAÇÃO EXTERNA EM PORTUGAL

Helder Sousa¹

Nota introdutória

Avaliar o impacto da avaliação externa num dado sistema educativo constitui uma tarefa de enorme complexidade, sobretudo se o enfoque for a melhoria do desempenho dos alunos, elemento central na avaliação do próprio sistema. Em Portugal, essa tarefa pode estar a ser dificultada, entre outros aspetos, pelo facto de todo o sistema de avaliação estar alicerçado em provas públicas e na incontornável necessidade de todos os anos produzir novos instrumentos de avaliação externa.

Medir a qualidade de um sistema educativo tendo por referência o papel e os resultados dos exames, tendo em conta o desempenho escolar dos alunos, constitui uma tarefa indissociável do comportamento de inúmeros atores. De forma mais fina, indicadores como taxas de retenção ou de abandono escolar ou ainda o desvio entre média de idades real de uma dada coorte de alunos e a idade modal para um dado ano letivo são cruciais na avaliação do sistema e são associadas, ainda que indiretamente, aos níveis de sucesso nessa mesma avaliação externa.

No caso nacional, todavia, aquela associação requer particular cautela. Uma análise realista e válida dos resultados da avaliação externa como contributo para essa métrica de avaliação do sistema requer uma contextualização de múltiplas variáveis que, em regra, não é tida em consideração.

Como ponto prévio, não é possível ignorar que a avaliação externa surge no percurso escolar dos alunos de forma pontual, no tempo – no final de cada ciclo – bem como no elenco de disciplinas frequentadas: duas disciplinas – Português e Matemática – nos três primeiros momentos de aplicação de provas externas, final do 1.º, do 2.º e do 3.º ciclos do ensino

¹ Presidente do Conselho Diretivo do Instituto de Avaliação Educativa, I.P..

básico; quatro disciplinas, em que o Português é transversal ao universo de alunos do ensino regular que pretendem concluir o ensino secundário, e outras três, escolhidas de um elenco de 19 em função do percurso escolar adotado.

Em segundo lugar, o contributo do resultado dos exames para a classificação final de um aluno interno está limitado a 30% do valor final. Os restantes 70% decorrem da avaliação interna.

Um terceiro aspeto prende-se com a avaliação dos alunos ao longo do seu percurso escolar regular: um aluno que nunca tenha sido retido é avaliado em sede de classificação final de frequência em cerca de 80 ocasiões, considerando o número de disciplinas e os anos de escolaridade frequentados. Mas, até concluir a escolaridade obrigatória, só realiza obrigatoriamente dez provas nacionais em sede de avaliação externa.

Dito isto, fica claro que os exames, em sentido estrito, e independentemente de se discutir o seu elenco, em termos das disciplinas avaliadas ou dos momentos, mais ou menos precoces no percurso escolar dos alunos em que são aplicados, constituem apenas uma pequena parte do problema.

Um contributo para uma reflexão coerente e sustentada do impacto dos exames na avaliação do desempenho do sistema educativo pressupõe, entre outros aspetos, ter em consideração a finalidade (ou finalidades) da sua aplicação. Afinal, para que fazemos exames? Para:

- certificar aprendizagens;
- seleccionar os alunos, no caso do ensino secundário;
- regular o ensino e a aprendizagem;
- monitorizar, ao longo do tempo, a qualidade do desempenho do sistema educativo.

Os exames, no caso nacional e em termos concretos, servem apenas uma, duas ou para múltiplas daquelas finalidades?

Além das finalidades formais da avaliação externa, consagradas de longa data no nosso quadro legislativo – certificação das aprendizagens e seleção para o ingresso no ensino superior – numa lógica de eficácia e

sobretudo de eficiência, considerando que fazer e aplicar exames tem custos que não são irrelevantes, é fundamental que se procure potenciar o papel dos exames também ao nível da regulação do ensino e da aprendizagem, bem como da monitorização diacrónica da qualidade do desempenho do sistema educativo.

A recorrente associação entre resultados de exames, práticas pedagógicas e qualidade do desempenho do sistema educativo é a dimensão que mais relevância assume no plano mediático e académico em Portugal, bem como na maioria dos países em que se aplicam este tipo de provas.

É, contudo, indispensável equacionar que tão grande ambição no aproveitamento dos resultados dos exames implica, para a concretização de uma análise correta e consequente dos mesmos, a consciencialização das limitações associadas ao sistema nacional de exames.

As considerações a seguir apresentadas propõem tão-somente uma apreciação superficial dos problemas e visam, principalmente, provocar o que se considera ser uma necessária e urgente reflexão sobre a problemática dos exames no nosso sistema de ensino.

A regulação do ensino e da aprendizagem e a comparabilidade dos resultados: limitações do sistema de avaliação nacional

O sistema de avaliação externa nacional, quer pela forma como está concebido, quer pela forma como parece ser vivenciado, principalmente por alunos e professores, e a avaliar pelos resultados observados nos últimos cinco anos, não mostra sinais de estar a contribuir para o que se pode considerar uma regulação positiva do ensino e da aprendizagem. Ou seja, os exames e o uso dos seus resultados deviam contribuir para, conjuntamente, informar alunos e professores dos domínios, na aprendizagem e na lecionação, em que são evidentes mais dificuldades.

Assim, de forma gradual, mas sustentada, deveria ser evidente para sucessivas coortes de alunos um desejável progresso traduzido na melhoria dos resultados, situação que, salvaguardadas variações de resultados que decorrem apenas de aspetos de ordem estatística, se deveria

traduzir num aumento geral das classificações médias dos exames nacionais, o que não se observa.

Como atrás se referiu, o sistema de avaliação externa nacional engloba um número de provas que se pode considerar adequado e os seus resultados têm um peso reduzido no cômputo da avaliação final em cada ciclo de estudos. Apresenta características específicas, ao nível da conceção das provas e no modo como são lidos e usados os resultados que, embora com forte potencial para uma eficaz regulação positiva do ensino e de promoção da melhoria das aprendizagens, não parecem ajudar à prossecução do papel de regulação a que atrás se aludiu, e que se justificaria plenamente, dado o investimento que a sua aplicação acarreta.

Como contributo para a reflexão sobre o papel dos exames no sistema de ensino, vale a pena destacar, entre outros, os seguintes aspetos:

- o facto de todas as provas de avaliação externa de alunos concebidas por instituições nacionais e aplicadas em contexto de exame serem públicas, ou seja, ano após ano é necessário a total reposição dos itens no sistema¹;
- a valorização, em termos académicos e mediáticos, dos resultados globais (indicadores macro associados essencialmente às médias nacionais), em detrimentos dos resultados por item, por domínio ou por tema, sempre que disponíveis;
- a disponibilização de informação (resultados) muito detalhada e com forte potencial para informar e apoiar decisões pedagógicas e opções didáticas de qualidade.

Implicações do carácter público das provas

i) Impactos positivos e negativos no funcionamento do sistema de ensino

A natureza pública das provas apresenta vantagens e desvantagens no que se refere ao impacto potencialmente positivo ou negativo no

¹ Considerando o número de provas finais de ciclo, exames finais nacionais e testes intermédios aplicados e produzidos anualmente, estima-se que, em média, sejam concebidos de novo, todos os anos, cerca de 2500 itens, dos quais cerca de 1800 se tornam públicos.

funcionamento do sistema de ensino, em geral, e de avaliação, em particular.

O conhecimento das provas e dos respetivos critérios de classificação por todos os interessados, logo após a sua realização, facilita a regulação dos processos de reapreciação, quando existem. Pode constituir também uma mais-valia para professores e alunos, que assim podem, em anos subsequentes, estar mais familiarizados com a natureza específica das provas para que preparam os alunos e com a tipologia e especificidade dos itens a que estes terão de responder.

Todavia, aquela suposta mais-valia comporta também aspetos reconhecidos como negativos: incentiva o chamado *training to the test*, que, em última instância, mais não é do que um treino que contamina de forma menos positiva o processo educativo na vertente processual inerente a uma correta forma de ensinar e de aprender: aprender para «ficar a saber» e não apenas para obter um resultado.

Aprende-se a técnica de responder às questões, não raras vezes em detrimento de uma aprendizagem focada no rigor dos conceitos, na precisão da linguagem e, principalmente, na capacidade de desenvolver operações mentais que sustentam e estruturam o processo cognitivo, consolidam e enriquecem o saber, tendo em vista uma aprendizagem que se quer para a vida e não para o imediatismo da obtenção de uma «nota» ou de um sucesso efêmero.

O conhecimento de uma extensa bateria de itens, disponível na página eletrónica do IAVE, a que se juntam as inúmeras publicações com mais umas largas centenas de itens de cada disciplina, pode estar a fomentar, de forma muito generalizada, um «estudo» que, estando centrado no treino atrás referido, legitima uma suspeição, embora não confirmada psicometricamente, de que o prévio conhecimento das provas poderá contribuir para inflacionar de forma artificial os resultados, ou pelo menos para sustentar a sua manutenção, sem oscilações estatisticamente relevantes, ao longo dos últimos anos.

Aquela tentação para o «treino», apenas pelo treino (embora haja também aspetos positivos em pensar e elaborar respostas a exames como estratégia de aprendizagem), é uma armadilha a que muitas vezes os alunos não conseguem fugir. A pressão social sobre os resultados, os *rankings*, ainda muitas vezes concebidos e divulgados sem qualquer contextualização de ordem socioeconómica, ou a habitual pressão mediática na análise dos resultados constituem os ingredientes perfeitos para que os exames não se estejam a traduzir num real e efetivo benefício para a melhoria das aprendizagens.

Entenda-se, para que não restem dúvidas, que o problema não reside nos exames enquanto instrumentos de medida do sucesso educativo, mas sim na forma como são vivenciados por quem os realiza e por quem divulga e por quem lê e usa os respetivos resultados. De acordo com a literatura especializada¹, a conceção de avaliação marcada pelos efeitos da sua componente sumativa está presente no sistema muito para além da realização dos exames: em contexto de avaliação interna, o peso relativo dos resultados dos testes no balanço final das classificações é muito significativo, mesmo em disciplinas e em anos de escolaridade que nunca estiveram sujeitos a provas finais ou a exames nacionais.

ii) Implicações financeiras e de mobilização de recursos humanos na manutenção do sistema de exames assente em provas públicas

Complementarmente aos aspetos de ordem pedagógica, vale a pena também equacionar a vertente financeira. A constante e ininterrupta produção de itens obriga a uma mobilização de recursos humanos muito superior à necessária numa situação de aplicação de provas não públicas, como, por exemplo, as provas aplicadas nos estudos internacionais de avaliação de alunos – PISA, PIRLS ou TIMSS – ou como o recentemente aplicado teste de Inglês *Key for Schools*.

¹ No âmbito do projeto PROMED, desenvolvido pelo ex-GAVE, foi possível recolher informação sobre práticas de avaliação interna num universo em mais de 100 escolas. As evidências das práticas descritas corroboram as considerações aqui referidas.

Uma opção pela utilização de provas não públicas permitiria, a prazo, uma redução muito substantiva dos custos associados à disponibilização de instrumentos de avaliação externa. Ultrapassada a fase de investimento na constituição de um banco de itens robusto, integrando um número de itens com diferentes propriedades e capazes de avaliar globalmente as aprendizagens decorrentes do currículo, a possibilidade de reutilização dos mesmos itens em avaliações consecutivas permitiria reduzir em mais de 80% o número de itens que atualmente é necessário produzir.

Aquela opção, em cuja aplicação não se tem investido, em grande medida pelos constrangimentos legais que se observam no que se refere à expectativa e direito dos destinatários em aceder aos enunciados, permitiria a disponibilização de itens previamente calibrados e a aplicação diacrónica de instrumentos cuja comparabilidade ficaria acima de qualquer dúvida: nestas circunstâncias, as flutuações de resultados deixariam de fora a variável associada à aplicação de provas que, embora concebidas no sentido de serem o mais similares possível, não deixam de ser todos os anos diferentes.

Naquele caso, poderia assumir-se que, excluídas as variáveis de contexto, cuja interferência nos resultados é sempre residual, e excluindo a normal variabilidade de resultados de natureza meramente estatística, as possíveis oscilações de resultados, positivas ou negativas, seriam um reflexo muito fino da qualidade do desempenho dos respondentes.

Também o esforço requerido em matéria de mobilização de recursos humanos seria naturalmente muito diferente. Os autores e outros especialistas envolvidos na conceção e validação técnica e científica das provas dedicam, nas atuais circunstâncias, a esmagadora maioria do tempo de colaboração com o IAVE à escrita dos itens e dos respetivos critérios de classificação, em detrimento de um maior envolvimento na análise dos resultados, no estudo do seu impacto no sistema ou na proposta de opções didáticas e pedagógicas que convirjam para uma sustentada melhoria da aprendizagem.

Sabendo-se que em Portugal a formação inicial dos docentes descursa, quase totalmente, a área da avaliação, e muito em concreto a área específica da conceção e validação de instrumentos de avaliação, a mobilização e a manutenção de equipas constitui, não raras vezes, um desafio difícil de ultrapassar, e que na solução atrás referida ficaria muito facilitada.

Sem um estudo exaustivo que permita aferir os eventuais ganhos financeiros na opção pela aplicação de provas não públicas, a avaliar pelos encargos com os custos atuais, e ponderando, em alternativa, um modelo de avaliação baseado em provas não públicas, pode-se estimar uma redução de custos da ordem dos 50% a 70% face aos valores atuais.

iii) Implicações da existência de provas públicas na chamada «comparabilidade dos exames»

A necessidade de todos os anos renovar os itens das provas de avaliação externa coloca um outro problema adicional, que se prende com as limitações técnicas na previsão e na validação dos resultados. Esta limitação constitui um dos principais constrangimentos na procura da tão pretendida comparabilidade das provas.

Em boa verdade, o que é feito, predominantemente em termos mediáticos ou por entidades representativas de professores das diversas disciplinas, é uma comparação *a posteriori* dos resultados observados em dois ou mais anos consecutivos. Se estes são idênticos infere-se que «as provas são comparáveis»; se diferem, de um ano para o outro, estando longe de ser consensual qual a variação da média do exame que sustenta a afirmação de que são diferentes, afirma-se que «as provas não são comparáveis».

Aquele discurso, fortemente enraizado na nossa sociedade, peca por uma enorme imprecisão técnica e concetual. Duas ou mais provas são comparáveis sempre e independentemente dos resultados que possam gerar. Mas o que de facto se pretende é que as provas aplicadas em diferentes ocasiões sejam estruturalmente semelhantes, ou seja, se esteja em presença do que se pode designar por «provas paralelas». O que equivale a dizer que se devia ter em conta, o que raramente acontece, a

consistência dos chamados moderadores de dificuldade¹, como sejam, entre outros, o conteúdo testado, os estímulos utilizados na formulação das perguntas, as características das tarefas propostas ou a consistência da resposta esperada, implicando uma adequação entre a pergunta e os respetivos critérios de classificação.

Ora não se conhece, até esta data, comentários públicos sobre provas passadas que explicitem uma análise prévia alicerçada nas características das provas, como atrás referido, pelo que, em regra, os comentários produzidos resultam de um olhar holístico, apriorístico e não raras vezes apressado. Esta urgência na pronúncia pública sobre as provas, muitas vezes resultado das exigências editoriais dos meios de comunicação social, não tem em conta as dimensões técnicas que constituem os elementos que permitem aferir do grau de semelhança ou dissemelhança, em termos de dificuldade, entenda-se, de duas ou mais provas.

Assim, no contexto nacional, em que falamos sempre de provas públicas, uma análise conducente a uma avaliação comparada do grau de dificuldade de provas terá sempre de ter em consideração:

- o objeto de avaliação e a sua relevância face ao currículo prescrito;
- o número de itens, total e por tipologia;
- a distribuição do número de itens por grau de dificuldade estimado²;
- a tipologia, o número e a complexidade dos suportes dos itens (textos, tabelas, gráficos, mapas, etc.);
- a distribuição das cotações pelos diversos itens;

¹ Leong, C. (2006). «On varying the difficulty of test items». *32nd Annual Conference of the International Association for Educational Assessment*. Singapura.

² O processo de determinação/previsão do resultado de cada item, o que no limite implica pensar na previsão do resultado global da prova, em provas públicas dificilmente pode alicerçar-se em procedimentos de pré-teste e de conseqüente validação das propriedades psicométricas do item – índice de dificuldade e de discriminação – por recurso, por exemplo, a técnicas estatísticas como o *Rasch Model* ou a *IRT* (Teoria de Resposta ao Item), entre outros, soluções amplamente aplicadas em provas não públicas.

- os critérios de classificação e os esquemas de classificação (pontuação de etapas de resolução ou de descritores de desempenho).

A comparação dos resultados das provas, de facto, não permite aferir, no imediato, se as estas são ou não equivalentes no seu grau de dificuldade. Comparar resultados constitui uma tarefa que se esgota na métrica adotada: diferença algébrica, variação percentual entre os resultados das provas de uma mesma disciplina. O valor obtido pode elucidar sobre situações totalmente distintas:

- resultados similares¹ podem ser registados em provas profundamente diferentes nas propriedades referidas;
- resultados com significativa disparidade podem ser registados em provas com elevado grau de semelhança, mais uma vez no que se refere às suas propriedades técnicas e estruturais.

Caso a caracterização pública de uma dada prova fosse precedida de uma análise técnica das suas propriedades, e daí se constatasse a existência de um elevado grau semelhança com uma outra prova tida como referência, aplicada no ano ou em anos letivos anteriores, seria possível inferir que a variação de resultados observada mediria, com elevado grau de confiança, uma dada evolução da qualidade do desempenho académico do universo de alunos que a realizou, por comparação com um outro universo, que se reporta ao da coorte de alunos que terá realizado prova equivalente em ano ou anos anteriores.

Ora, na falta dessa análise sobram referências ao aparente grau de dificuldade de uma dada prova logo após a sua aplicação. Não raras vezes são as opiniões de alunos ou de professores à saída de uma dada escola, noutros casos são as opiniões das associações de professores representativas das disciplinas envolvidas que aferem a suposta maior ou menor dificuldade da prova. A comunicação social tem, por regra, o papel

¹ Não existe um valor, absoluto ou relativo, sobre o qual haja consenso ou qualquer baliza tecnicamente válida quanto ao que se entende por resultado similar. Deste modo, a medida de semelhança reveste-se, em si mesma, de uma profunda limitação operacional.

de ampliar essas opiniões, num momento em que ainda estamos longe de poder conhecer quaisquer resultados.

O valor daquelas opiniões é muito relativo. Ano após ano assiste-se a uma frequente inconsistência entre as opiniões expressas no imediatismo do conhecimento do enunciado e os resultados mais tarde divulgados: coincidências entre afirmações de maior facilidade ou de maior dificuldade e a variação positiva ou negativa dos resultados de um ano para o outro são tão comuns como o seu contrário e, mesmo quando coincidentes, as variações, pela sua diminuta amplitude, não têm expressão estatisticamente relevante.

Ainda que possam existir antevisões de subidas ou descidas de resultados que se revelem acertadas, a verdade é que as variações interanuais dos resultados – considerando a média nacional dos exames, o indicador mais utilizado – quer nas provas do ensino básico, quer do ensino secundário, têm apresentado, por regra, amplitudes dentro do previsto, estatisticamente expectáveis e não significativas, para este tipo de variável: inferiores a 10 pontos no ensino básico e inferiores a 15 pontos no ensino secundário¹.

Aqueles valores de referência, cuja determinação se baseia numa análise de longo prazo das variações interanuais de resultados de exames nas disciplinas do ensino básico e secundário, apenas constitui uma baliza operacional que a todo o momento pode ser revista e questionada. Todavia, e independentemente de outros valores a ter em consideração, nunca se poderá esquecer que:

- a) o resultado final de uma prova constitui o somatório dos resultados obtidos nos itens que a compõem;
- b) o modelo de conceção e aplicação de provas vigente replica anualmente a disponibilização de provas cuja estrutura e constituição, no que se refere à tipologia dos itens e, conseqüentemente, também à natureza das operações mentais que

¹ Valores tidos por referência no trabalho desenvolvido pelo IAVE e que foram utilizados como critérios na avaliação do ex-GAVE no âmbito do QUAR.

estão subjacentes a uma adequada resolução dos mesmos, se têm mantido constantes¹;

- c) variações mínimas nos resultados de itens similares publicados em provas de anos sequenciais (admita-se que apenas de 1, 2 ou 3 pontos, em 200), amplificadas pelo número de itens das provas (em muitos casos da ordem das duas dezenas ou mais) podem gerar variações totais cuja amplitude é fácil de antever;
- d) variações de resultados cuja ordem de grandeza se situa abaixo de 10% da amplitude da escala de pontuação utilizada nas provas (100 ou 200 pontos, consoante os níveis de ensino) carecem de uma análise técnica e estatística detalhada até que se possa inferir, com um mínimo de rigor, sobre a natureza das variáveis explicativas para tal variação: apenas um facto de natureza estatística, que traduz o normal comportamento de valores gerados no contexto em que o são; ou, de facto, uma de duas situações – uma prova concetualmente mais fácil ou mais difícil que a do ano, ou anos, anteriores, ou, por sua vez, um sinal de que a coorte avaliada num dado ano evidencia uma prestação cuja qualidade global é melhor ou pior que as suas antecedentes.

Aqueles aspetos raramente são tidos em conta nos juízos de valor e nas múltiplas declarações que são produzidos acerca dos resultados dos exames e do seu impacto no sistema de ensino.

Sem esgotar a questão, é por isso necessário refletir sobre o interesse e a potencial mais-valia do contínuo e repetido uso mediático da informação produzida em torno da aplicação dos exames. Trata-se de um processo que, acima de tudo, alimenta o interesse de alguma comunicação social, fortemente marcado no tempo (decorre, em regra, ao longo de três meses, de maio a julho), com forte impacto social e político, mas que, ano após ano, se esgota no final de cada verão, ficando óbvio o seu reduzido

¹ As alterações de estrutura de uma dada prova apenas têm ocorrido em duas circunstâncias: em consequência de alterações curriculares; por evidente obsolescência ou inadequação de opções em matéria de conceção da prova ou de itens, o que, em regra, gera atualizações desenvolvidas em estreita colaboração com os membros do Conselho Científico que integram o IAVE (ex-Conselho Consultivo do GAVE), onde estão representadas as associações de professores e sociedades científicas das várias áreas disciplinares.

interesse em procurar informação relevante para uma discussão tecnicamente válida e consistente sobre o impacto da avaliação externa no sistema educativo.

As leituras de resultados

Para além da mediatização das opiniões apriorísticas sobre o grau de dificuldade das provas, exploradas no dia da sua realização ou nos dias imediatos, o tipo de análises que enchem as páginas dos jornais ou alimentam debates televisivos tomam por referência leituras de resultados que apenas ajudam à especulação e a uma certa agitação social.

Alguns exemplos são esclarecedores dos objetivos que muitas vezes subjazem às notícias publicadas.

O uso das classificações totais na divulgação dos resultados dos exames do ensino secundário, em detrimento das classificações dos alunos internos, é prática recorrente. Este indicador, que resulta do cálculo da média dos resultados do total de alunos, internos e autopropostos, revela-se inadequado para caracterizar o desempenho da população de alunos em análise, pelo que a sua divulgação é desprovida de sentido. Os alunos internos são os que completaram o seu percurso escolar regular, frequentaram as aulas até ao final do ano letivo e obtiveram uma classificação final que legitimou a sua presença no exame: ou seja, tiveram o aval dos respetivos professores para a realização do exame final. Os alunos autopropostos, ou a sua larga maioria, a avaliar pelos resultados, são os que, antecipando a sua reprovação em sede de avaliação interna, optam por anular a matrícula: muito provavelmente, não teriam, no final do ano letivo, o aval dos seus professores para a realização do exame.

A diferença entre a média de um e de outro grupo de alunos, internos e autopropostos, assume valores muito desiguais de disciplina para disciplina, bem como o peso relativo dos alunos autopropostos no universo dos alunos que realiza a prova, pelo que é desprovido de significado a sua utilização no contexto em que é feita. Aquela diferença chega a ser superior a 4 valores (40 pontos em 200). Como se sabe,

agregar os resultados de populações com características tão diferentes é mesmo um exemplo clássico, no ensino da estatística, de situações em que não se deve utilizar uma média simples.

Todavia, para servir intuitos mediáticos, como o valor da média total é, pelas razões apontadas, sempre inferior ao valor da média dos alunos internos, acaba por ser o mais utilizado por diversos órgãos de comunicação social.

Outra situação muito frequente é a contabilização do número de disciplinas cujos resultados são inferiores ou superiores aos observados no ano imediatamente anterior. Trata-se de mais um exercício sem qualquer sentido. Em primeiro lugar, porque no conjunto das 23 provas do ensino secundário haverá sempre um determinado número de subidas e de descidas, que é variável de ano para ano. De facto, o que é menos provável é que numa mesma disciplina a variação entre dois anos consecutivos seja igual a zero. Em segundo lugar, porque a simples contabilização do número de disciplinas cujos resultados sobem ou descem (o enfoque é maior quando o número de descidas supera o das subidas) não tem em consideração a amplitude dessas variações. Fica à consideração do leitor imaginar como seriam capas de jornais se, por um qualquer imprevisto estatístico, num dado ano os resultados descessem em todas as disciplinas ainda que apenas um ou dois pontos (em 200).

Como se referiu, as leituras de resultados que mais frequentemente se divulgam carecem de maior reflexão e aprofundamento para que delas se possam extrair informações relevantes e significativas para uma adequada avaliação do sistema de ensino. De entre outros exemplos que se poderiam seleccionar, fica uma chamada de atenção para o uso descontextualizado da diferença de resultados entre a classificação interna de frequência (CIF) e a classificação do exame (CE).

No que se refere à relação entre as classificações de exame (CE) e as classificações internas de frequência (CIF), pode realçar-se a manutenção de uma correlação positiva e relativamente forte entre as duas classificações (CE e CIF) nas disciplinas com maior afluência de alunos,

como são os casos das disciplinas como Física e Química A, Matemática A e Biologia e Geologia (entre 0,77 e 0,79). Este indicador é ligeiramente inferior aos anteriores em disciplinas da área das ciências sociais e humanas, como Geografia A, História A, Economia A ou Filosofia (entre 0,59 e 0,71) e na disciplina de Português (0,64). Com estes valores percebe-se um elevado grau de correspondência entre os desempenhos de uma parte significativa dos alunos em sede de avaliação interna e externa.

Fica, no entanto, por explicar a diferença, em valores (ou pontos) entre aquelas duas classificações obtidas a partir de instrumentos e processos distintos de avaliação. Todavia, é especulativo afirmar que a validade dos exames, ou dos seus resultados, possa ser posta em causa pelo facto de nas disciplinas com piores resultados nos exames (por exemplo, Matemática A ou Física Química A) as diferenças entre a CE e a CIF serem muito acentuadas. De facto, se atendermos às significativas diferenças no objeto de avaliação, que constitui o suporte para a produção de resultados em ambos os casos, ou às componentes que concorrem para a composição das classificações (muito mais diversificadas em contexto de avaliação de sala de aula), as diferenças, sistematicamente observadas em todas as disciplinas, podem ser compreensíveis e não devem justificar, sem uma análise mais profunda, quaisquer outras leituras.

Mais difícil de entender é a significativa regularidade da CIF, considerando as 11 disciplinas com mais de 2500 alunos internos sujeitos a exame. Com exceção de Desenho A, Geometria Descritiva A ou Economia A (com CIF entre 14,2 e 15,2 valores), todas as outras, onde se incluem disciplinas tão díspares no que se refere à sua natureza, públicos-alvo e que desde sempre têm apresentado resultados da avaliação externa tão divergentes (Matemática A, Física Química A por oposição a Português, Geografia ou Filosofia), apresentam um valor médio da CIF que se confina ao intervalo entre 13,0 e 13,7 valores. Esta circunstância, aliás uma marca crónica do nosso sistema de ensino, pode, outrossim, levar-nos a questionar a validade dos resultados atrás referidos, que não espelham a reconhecida especificidade e diversidade das disciplinas em

apreço, no que se refere, entre outros aspetos, às conceções sociais sobre a complexidade das respetivas áreas científicas.

Uma última nota para a utilização dos resultados dos exames como medida para o estabelecimento de rankings das escolas sem contexto social de suporte. Trata-se de uma situação que, embora tendencialmente se esteja a corrigir, ainda introduz uma profunda distorção na forma como se processa a apropriação social da qualidade do desempenho das organizações escolares.

De facto, qualquer valorização da qualidade da prestação apenas centrada nos resultados dos exames, baseada em resultados médios globais, deixa de fora elementos de análise essenciais. Como é reconhecido internacionalmente, os indicadores de natureza socioeconómica que contextualizam a situação de partida dos alunos envolvidos são indissociáveis de uma avaliação credível e justa da qualidade da prestação educativa das escolas em apreço. Numa escola que se pretende inclusiva e capaz de minimizar os efeitos da retenção ou da exclusão, um indicador da progressão da qualidade da aprendizagem dos alunos envolvidos no ciclo de estudos que culmina com o exame ou a medida da percentagem de alunos à entrada e à saída desse mesmo ciclo são também elementos cuja análise se considera essencial e ainda por fazer no nosso sistema.

A divulgação dos resultados desagregados por item: um potencial ainda por explorar

Como atrás se referiu, uma das vantagens em dispor de provas públicas consiste no conhecimento sem restrições de todo o manancial de itens e dos respetivos critérios de classificação, por oposição a um sistema de provas não públicas em que apenas é habitual a libertação de alguns itens que constituem exemplos ilustrativos das provas publicadas.

A conjugação do conhecimento de todos os itens publicados com os respetivos resultados constitui um recurso que não é vulgar encontrar disponível em outros sistemas educativos.

No sistema nacional, desde 2007, primeiramente no contexto das antigas provas de aferição do ensino básico e depois numa prática alargada a todas as provas, assiste-se a uma combinação de oferta de resultados que junta informação macro, como por exemplo as médias nacionais ou a percentagem de classificações «positivas», divulgada quer pelo JNE, quer pelo IAVE, disponível para o público em geral, com informação detalhada ao nível do item, desde há vários anos divulgada pelo ex-GAVE e agora pelo IAVE.

Os resultados por item estão a ser progressivamente disponibilizados ao público em geral, embora cada escola receba anualmente um relatório, de acesso restrito, com os seus próprios resultados, por item, por turma, por escola (escolas integradas em unidades orgânicas ao nível do agrupamento).

Assim, presentemente, temos disponível informação sobre os seguintes indicadores:

- Combinação de resultados ao nível macro e micro com informação georreferenciada, desde 2008;
- Resultados por item e por escola a nível regional e nacional;
- Possibilidade de situar os resultados da escola, por disciplina e por item, no contexto regional e nacional;
- Possibilidade de identificar na escola as áreas em que os alunos evidenciam fragilidades em cada disciplina.

Esta informação, que carece de uma leitura e interpretação por parte dos professores de cada disciplina e dos órgãos de gestão de cada unidade orgânica, apresenta condições para propiciar um «*feedback* diferido» com implicações na reformulação das estratégias de ensino e de aprendizagem.

Não se trata, como se por vezes implícito em alguns escritos sobre a matéria, em dar *feedback* aos alunos que realizaram um dado exame, mas aproveitar o que a sua prestação nos ensina para aplicar nos anos subsequentes as medidas pedagógicas e didáticas que permitam corrigir ou minorar as limitações observadas nos alunos que irão realizar exames nos anos seguintes.

Uma análise sistemática de resultados por item ao longo de séries com cinco ou mais anos, tendo por referência itens com elevado grau de semelhança, como tem sido a prática nas provas nacionais, não mostra, em termos gerais, que aquele potencial *feedback* nas práticas de ensino esteja a propiciar melhorias visíveis e sustentadas na qualidade das aprendizagens realizadas pelos alunos.

A evolução dos resultados e o papel dos exames

A evolução mais recente dos resultados mostra-nos que numa série mais longa, superior a cinco anos, independentemente das variações anuais, não há uma tendência clara e sustentada de subida ou descida das médias nacionais, da percentagem de alunos com classificações ditas «negativas» ou «positivas», e mesmo os gráficos de frequências, que ilustram as distribuições dos resultados, apresentam uma configuração muito similar.

O conhecimento de um largo acervo de itens disponível para consulta pública, bem como a manutenção de provas cuja conceção observa princípios que visam preservar a sua semelhança, são condições que permitem uma familiarização dos alunos com as provas que realizam assim como um «treino» para realização dos exames operado com itens muito similares aos que acabam por encontrar nas suas provas. Estas circunstâncias poderiam até sustentar e explicar uma expetável subida dos resultados, o que está longe de se observar.

Ora aquela informação (detalhada), cruzada ou não com os relatórios nacionais, permite, sempre que se proceda a uma análise fina das evidências que esses resultados ilustram, detetar para cada escola as áreas específicas de aprendizagem que cronicamente mostram ser aquelas em que um número significativo de alunos revela mais dificuldades.

Assumindo que, em termos médios, as sucessivas coortes de alunos registam percursos escolares que antecedem o momento da realização de um dado exame marcados por processos de aprendizagem que geram níveis de proficiência similares, toda a informação que em resultado de cada exame é possível compilar deveria conduzir à progressiva implementação de intervenções didáticas especificamente focadas na

tentativa de superação das reconhecidas dificuldades. No entanto, como foi referido, mais uma vez a informação de que se dispõe não mostra sinais sustentados de melhoria nos resultados de itens similares na generalidade das disciplinas, ou seja, itens que avaliam, ao longo de anos sucessivos, os mesmos conteúdos ou as mesmas operações mentais.

No atual contexto, os exames constituem uma ferramenta indispensável para assegurar um referencial nacional único na certificação da qualidade da aprendizagem dos alunos, e nessa medida, podem também constituir um instrumento de medida útil na seleção de alunos, nomeadamente no acesso ao ensino superior¹.

Todavia, não esqueçamos que o papel dos exames não se deve esgotar naquelas duas finalidades. Num quadro de eficiência decorrente da aplicação dos exames é essencial usar positivamente o seu potencial efeito de regulação do ensino e da aprendizagem e de monitorização da qualidade do desempenho do sistema educativo. Ainda que aquelas finalidades não pressuponham, em si mesmas, juízos de valor quanto a aspetos de ordem qualitativa, aquilo que se espera de um qualquer sistema social, de que a educação não é exceção, é uma progressiva melhoria.

No plano da regulação da qualidade do ensino e da aprendizagem, parece ser notório que os exames têm sido boas ferramentas para sustentar a manutenção de um *status quo* marcado por duas evidências socialmente bem aceites: médias nacionais em torno do valor da linha de corte entre classificações ditas «negativas» e «positivas»; distribuições de frequência dos resultados em que cerca de 50% dos alunos alcançam classificações de exame inferiores ao valor da dita linha de corte.

Os exames podem e devem ser também uma ferramenta para uso formativo, o que, no caso nacional, fica particularmente facilitado pela conjugação de provas públicas e de uma política de divulgação de

¹ A discussão em torno da valia dos exames como instrumentos de seleção de alunos no acesso ao ensino superior não pode nem deve ignorar as limitações que decorrem do uso dos seus resultados. O recurso a eventuais opções alternativas de seriação dos candidatos mas, sobretudo, o uso de outros critérios subjacentes à forma de cálculo utilizada ou o uso de classificações mínimas são aspetos que deveriam merecer profunda reflexão e eventual alteração.

resultados que privilegia uma informação de elevada qualidade e detalhe. É nesse sentido que se entende que a regulação dos exames deveria ter um impacto positivo claro e sustentado na qualidade da aprendizagem, com expressão numa subida gradual dos resultados.

Assim, no que se refere à monitorização do sistema de ensino, aquilo que os resultados dos exames nos mostram, a avaliar pelos dados dos últimos cinco anos, sem prejuízo, como atrás referido, das limitações impostas pela natureza pública das provas, é a expressão de uma estagnação apenas quebrada, pontual e localizadamente em certos anos de escolaridade e disciplinas, por alguns indícios de melhoria que nos anos mais próximos será necessário confirmar.

Os resultados da avaliação externa: desafios e preocupações

Numa abordagem diacrónica e prospetiva é incontornável perspetivar a educação num horizonte de longo prazo e, nesse sentido, o que hoje já se sabe sobre os pontos fortes e fragilidades do nosso sistema de ensino deve sustentar reflexão que possa concorrer para a manutenção, melhoria ou alteração das opções que hoje se conhecem.

Entre outras situações, podemos considerar como principal e mais urgente desafio que se coloca ao sistema de ensino a necessidade de valorização da qualidade da aprendizagem, independentemente dos percursos académicos dos alunos.

Se quisermos atribuir aos exames algum papel naquele processo é essencial que a informação que deles se extrai contribua para corrigir, ao nível dos processos cognitivos e das operações mentais que constituem os alicerces do saber e do saber pensar, as fragilidades detetadas em cada momento.

Cada aluno, com os seus professores ou de forma autónoma, pode e deve usar os resultados dos seus exames (ou do treino para a sua realização) como uma ferramenta de autoavaliação e de autoaprendizagem. Já ao nível da escola, do departamento, do grupo disciplinar, ou mesmo a título individual, os resultados dos exames devem servir, não tanto para corrigir

os erros específicos detetados neste ou naquele aluno, mas para uma reflexão sobre as razões que subjazem à obtenção de resultados menos conseguidos num dado item ou conjunto de itens.

Mais do que olhar para o resultado médio da escola, processo que continua a dominar a atenção da maior parte dos intervenientes, é indispensável refletir e agir a partir dos resultados por item de uma dada prova na arquitetura e conceção do processo de ensino dos alunos a frequentar os anos letivos precedentes. Sem prejuízo de se poder encontrar esta ou formas similares de atuação em algumas escolas, ou em alguns contextos disciplinares específicos dentro de algumas escolas, estamos longe de poder constatar ser esta a forma mais habitual de usar a informação gerada pelos resultados dos exames.

A cristalização de resultados a que se tem assistido no plano da avaliação externa nacional constitui, assim, a mais relevante preocupação quanto à possível evolução do sistema de ensino medida a partir deste indicador. A avaliar pela forma como se lida com os resultados dos exames, parece assistir-se a uma aceitação acrítica de um sistema que não mostra sinais de evolução positiva, que se replica, numa consentida mediania, de forma mais ou menos imutável ao longo de sucessivas coortes de alunos.

Os exames, muitas vezes apontados como uma das causas do problema do insucesso ou da falta de qualidade da aprendizagem, pelo facto de condicionarem as práticas de ensino e de aprendizagem, têm um potencial formativo de cujo aproveitamento não existem evidências sistemáticas e numericamente expressivas. Se é certo que as preocupações com os resultados estão na ordem do dia, é também verdade que se continua a olhar para os resultados que menos ilustram e informam sobre a melhor forma de intervir na superação dos problemas de aprendizagem que persistem em afetar uma significativa percentagem dos nossos alunos: os resultados médios por escola ou globais, aqui designados por resultados macro.

Aqueles resultados, como se viu, condicionados pela vertente pública das provas, incapazes de devolver um diagnóstico fino sobre os problemas que

carecem de intervenção didática específica, têm sido sobrevalorizados em detrimento dos resultados micro, ao nível do item, cuja apropriação carece ainda de uma aprendizagem por fazer pela larga maioria dos professores, dos alunos e mesmo dos pais.

Parecem longe de estar esgotadas as potencialidades formativas de um adequado uso dos exames, quer na fase de preparação para a sua realização, quer, posteriormente, na análise dos seus resultados.

Investir nesta abordagem da avaliação deveria constituir um dos focos a privilegiar na formação de docentes. A aposta na generalização do uso da avaliação despida de conceitos estanques, ainda partida nas suas dimensões sumativa e formativa, mas antes como uma ferramenta capaz de ajudar a promover o sucesso, é o caminho que desde há quase duas décadas se procura traçar em sistemas educativos de outros países, onde há muito se percebeu que a solução não reside na ausência de avaliação formal e externa mas sim na procura de uma utilização verdadeiramente conjugada dos vários instrumentos de avaliação, aplicados interna e externamente, com mais ou menos enfoque formativo ou sumativo, na promoção de uma aprendizagem de qualidade.

No caso nacional, dispor de informação sistemática sobre os resultados por item constitui um recurso que permite mesmo ultrapassar um dos obstáculos concetuais que continua a alimentar a discussão em torno da viabilidade do uso formativo da avaliação externa.

Como preocupação final, um olhar sobre a chamada «linha de corte», o valor que, na escala de medida, se convencionou como a fronteira entre o sucesso e o insucesso. De facto, a tradição de considerar 50 pontos, na escala de 0 a 100 pontos, ou 95 pontos (100 pontos por arredondamento), na escala de 0 a 200 pontos, como as marcas que é necessário superar para alcançar a aprovação não deixa de ter efeitos perversos na forma como socialmente pensamos a aprendizagem. Aceita-se que o limiar de sucesso se alcance num contexto em que se aprendeu «metade» do esperado, como se aceitam como «normais» as distribuições de frequência de resultados

que apresentam uma configuração de curva normal, gaussiana, em que cerca de 50% dos alunos se situam abaixo da dita linha de corte.

Se é certo que a aceitação de níveis de exigência como os referidos não é universal no seio das famílias, é igualmente verdade que para muitas o principal desafio reside na obtenção de uma classificação «positiva», ainda que próxima da referida linha de corte.

Discutir esta conceção de avaliação e de medida seria um bom exercício coletivo. Como é sabido, falamos de uma convenção. Noutros países e noutros sistemas de avaliação entende-se que a linha de corte deve apresentar outros valores. A título de exemplo, no teste de inglês, *Key for Schools*, aplicado em Portugal no ano letivo passado, a classificação mínima que permite alcançar o nível para que o teste foi concebido¹ é 70 pontos em 100.

A questão é preocupante a dois níveis: globalmente, ao estarmos a legitimar a formação de jovens alicerçada na aceitação de um desempenho mediano como referência de qualidade, minimizando o desafio de considerar a aprendizagem um caminho sem fim; nomeadamente na formação e na avaliação das crianças no início do seu percurso escolar, em concreto ao longo do 1.º e do 2.º ciclo do ensino básico.

De facto, há muito que se reconhece que a aprovação naquela fase inicial e crucial da formação com resultados próximos de referida linha de corte indicia falhas graves na aprendizagem que se repercutem, e tendem a ampliar, nos níveis de sucesso em ciclos de ensino subsequentes.

Considerar outro referencial para a medida de aprovação não deve pressupor aceitar um aumento de retenções mas antes um incremento dos níveis de exigência na formação dos alunos, bem como um indispensável acompanhamento específico enquanto forem evidentes as lacunas que impeçam desempenhos com outra expressão quantitativa em sede de avaliação.

¹ Nível de proficiência linguística A2, de acordo com o Quadro Europeu Comum de Referência (QEQR).

Em termos concretos, por regra, os alunos que nos ciclos iniciais obtivessem uma classificação de nível 3, inferior a 70%, deveriam beneficiar de apoio extraordinário. Este apoio deveria ser desenhado em função das principais dificuldades de aprendizagem diagnosticadas quer em sede avaliação interna, quer pelas evidências dos resultados alcançados na avaliação externa.

A informação de que hoje os professores dispõem nestes níveis de ensino possibilita, pelo que atrás se referiu, um diagnóstico muito preciso das fragilidades de cada aluno, conhecida pela análise dos resultados por item ou pela informação agregada por tema ou por domínio¹.

Aquela circunstância permite apoios que podem ser pensados com elevada especificidade, isto é, não estamos a falar de apoios generalizados numa disciplina específica durante um ou mais anos letivos, mas antes de intervenções individualizadas e dedicadas à superação de dificuldades concretas, o que pode justificar a oferta de apoios educativos de duração não necessariamente longa mas antes pontuais, de curta ou média duração.

No plano de potenciais desafios, e apenas para citar os que se podem considerar essenciais e concretizáveis num quadro realista, tendo em conta os recursos disponíveis, podemos considerar a urgente modernização e atualização tecnológica de toda arquitetura associada ao processo de aplicação e classificação de provas, que, à semelhança do que hoje se observa noutros países, deverá ser sustentada por ferramentas que permitam a desmaterialização de todo ou de grande parte do processo. É exemplo a aplicação de testes em suporte digital e a respetiva classificação num mesmo suporte, processo que deverá igualmente evoluir para a aplicação, ainda que em circunstâncias específicas, de testes adaptativos (CAT).

¹ No caso da utilização dos resultados da avaliação externa como suporte da informação que visa caracterizar o perfil de aprendizagem de um dado aluno, é essencial ter em conta as limitações associadas ao facto de estarmos a considerar uma informação centrada numa bateria restrita de itens e por vezes num único instrumento de avaliação. As leituras que se devem fazer terão de ser enquadradas com informação mais consistente que a avaliação interna deve ser capaz de gerar.

Se associarmos aqueles processos à migração para um quadro de provas não públicas, situação que, embora exija um quadro legal específico que carece de algum desenvolvimento, se justifica plenamente, pelo menos nas provas aplicadas no ensino básico, estamos em condições de usufruir de vantagens em termos de qualidade e simplificação dos processos há muito conhecidas: forte redução de custos, pelo menos a médio prazo; maior objetividade e fiabilidade na classificação das provas; incremento da validade dos resultados, seja pela aplicação de itens submetidos a pré-teste e calibrados, seja pelo alargamento substancial da bateria de itens a aplicar.

Em suma, é necessário que pensar sobre os exames seja pensar no conjunto de pressupostos que lhes estão subjacentes e no conjunto de medidas educativas e processuais que podem contribuir para incrementar a sua eficácia e eficiência no sistema educativo.

UM OLHAR SOBRE OS DADOS DA AVALIAÇÃO EXTERNA DA APRENDIZAGEM

Luís Santos¹

I – Olhares sobre a avaliação externa

A avaliação externa da aprendizagem é considerada, pelo estudo PISA 2009, um fator crítico de sucesso para que qualquer sistema de ensino possa evoluir no que diz respeito à qualidade do processo de ensino e aprendizagem, sendo geralmente aceite que a sua implementação produz um efeito muito significativo sobre algumas variáveis do âmbito do trabalho desenvolvido pelas escolas, designadamente:

- Induz práticas inovadoras de ensino e de avaliação nas escolas
- Regula as avaliações internas
- Fornece dados para avaliação do sistema educativo, para tomada de decisões
- Regula o sistema — dá indicações às escolas, aos professores e aos alunos acerca do que é considerado importante ensinar e aprender

A avaliação da aprendizagem é geralmente considerada como um processo sistemático e padronizado de recolha de informação acerca dos conhecimentos e capacidades adquiridos pelos alunos, no âmbito do processo de ensino e de aprendizagem, e que permite a formulação de apreciações acerca do trabalho desenvolvido.

Estas apreciações, elaboradas por diversos atores e entidades mais ou menos conhecedores do contexto educacional, baseiam-se tipicamente em análises estatísticas, nem sempre esclarecedoras, realizadas sobre os dados gerados neste processo.

Pretende-se nesta comunicação apresentar outros olhares sobre os dados da avaliação externa, que possam servir de base para estudos mais

¹ Presidente do Júri Nacional de Exames.

profundos que contribuam decisivamente para uma melhor compreensão e organização do currículo.

A questão “O que fazer com os dados da avaliação externa?” leva-nos naturalmente, numa dinâmica analítica, a questionar se podemos ir mais além do que o cinzento obscuro das médias das classificações de exame, métrica a que infelizmente se resume o debate público sobre esta matéria. Esta pergunta suscita-nos uma série de outras questões através das quais podemos refletir, nomeadamente:

- Que abordagens e olhares sobre a avaliação externa da aprendizagem?
- Quais os dados estatísticos mais adequados para uma análise efetiva dos resultados da avaliação externa?
- A avaliação externa regula efetivamente os processos pedagógicos de ensino e aprendizagem?
- A avaliação externa permite-nos olhar de forma diferenciada para os vários tipos de alunos?
- A avaliação externa promove uma verdadeira autoavaliação nas escolas?
- O Ranking das escolas dá-nos uma real perceção da qualidade das escolas?

Podemos definir três níveis de análise relativamente aos dados da avaliação externa: o olhar dos decisores políticos, o olhar das escolas e o olhar do público.

Os decisores políticos devem ter uma abordagem aos dados da avaliação externa que lhes permita obter informações importantes e consistentes sobre o sistema educativo. Centralmente, os dados da avaliação externa são olhados, em geral, de um ponto de vista que permite habilitar o MEC para a tomada de decisões relativas, principalmente, aos seguintes aspetos:

- Aferição dos conhecimentos dos alunos a nível nacional
- Adequação dos currículos
- Adequação dos programas das disciplinas
- Regulação das práticas pedagógicas e das didáticas das disciplinas
- Adequação aos vários tipos de alunos e escolas
- Avaliação das escolas

O olhar das escolas sobre os dados da avaliação externa tem como objetivo um verdadeiro e efetivo processo de autoavaliação, pelo que deve ter em conta três níveis principais de análise, nomeadamente:

- Análise de nível macro – comparação com os dados estatísticos nacionais
- Análise de nível intermédio – comparação com os dados estatísticos regionais (NUTS III e Concelho)
- Análise de nível local – análise fina dos resultados obtidos pelos alunos da escola

O olhar do público é normalmente limitado aos *Rankings*, listas de dados apresentados pelos meios de comunicação social, os quais têm um processo de análise dos dados da avaliação externa que se resume, salvo raras exceções, à apresentação das médias das classificações obtidas pelos alunos de cada escola nas provas finais e nos exames nacionais.

II – *Que dados e métricas devemos utilizar para uma análise adequada da avaliação externa?*

Infelizmente a generalidade das análises efetuadas por todos os agentes que têm influência, interesse ou participação no processo educativo são totalmente monocórdicas no que diz respeito aos indicadores estatísticos utilizados, o que leva a que apenas sejam valorizadas as médias das classificações.

Com a utilização generalizada de um único indicador estatístico, podemos concluir que a sociedade portuguesa é, no mínimo, “*médio-dependente*”. Esta é uma situação problemática, já que a média é, como todos sabemos, uma medida estatística que tem o potencial de esconder um grande número de informações fundamentais para uma correta análise estatística dos dados da avaliação externa por todos os *stakeholders*, sejam os responsáveis políticos, as escolas ou o público em geral.

É, por conseguinte, muito importante que nas diversas análises dos dados das provas finais de ciclo ou dos exames nacionais sejam utilizados métodos e medidas estatísticas que possam complementar a única que é tipicamente utilizada. Assim, é muito importante utilizar uma maior

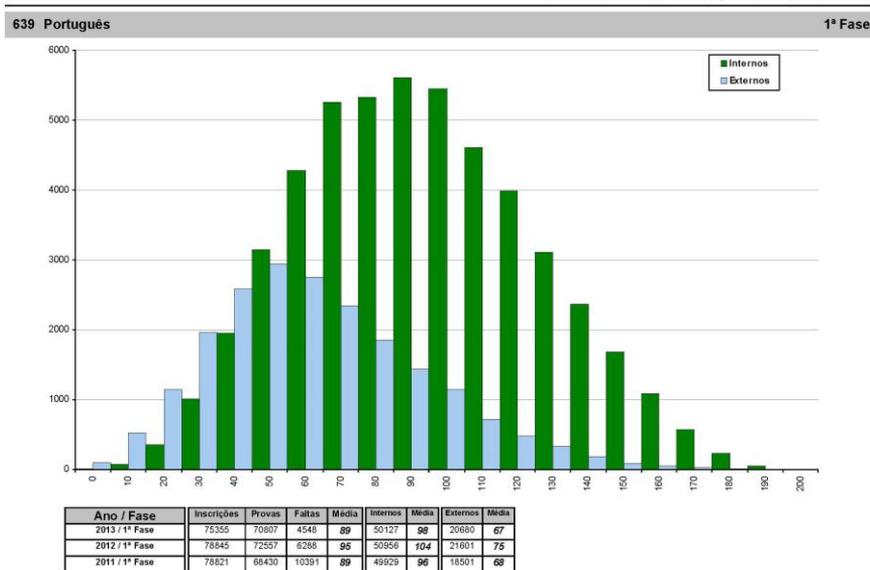
diversidade de medidas estatísticas que possam caracterizar de forma mais completa e fina as distribuições das classificações das provas finais e dos exames nacionais, nomeadamente, medianas, medidas de dispersão, desvio padrão e coeficiente de variação, etc. Para além desta maior diversidade de indicadores, é também muito importante a realização de análises regionais, que nos permitem perceber as variações de índole local e regional, bem como outro tipo de análises significativas para avaliar a qualidade dos processos de ensino aprendizagem nas escolas, designadamente a análise das diferenças entre as classificações internas (CIF) e as classificações de exame (CE).

Como exemplo do tipo de análise, podemos observar a distribuição das classificações do exame nacional de Português (639), da 1.^a fase de 2013. Como se pode verificar, trata-se de uma distribuição descrita por uma curva de Gauss, com uma baixa dispersão dos dados. Neste caso a média caracteriza de forma satisfatória a distribuição, pelo que ao dizermos que esta tem o valor de 89 pontos, a probabilidade de uma classificação se encontrar nas vizinhanças da média é muito elevada.

Por outro lado, é também muito importante fazer a desagregação dos dados entre alunos internos e autopropostos, já que estes dois grupos de alunos têm desempenhos bastante diferenciados, mais evidentes nas disciplinas trienais do que nas disciplinas bienais, como veremos mais à frente. Na distribuição em apreço é notória esta diferença, se observarmos as barras azuis, que representam a distribuição das classificações relativa aos alunos autopropostos.

EXAMES FINAIS NACIONAIS DO ENSINO SECUNDÁRIO 2013

Distribuições de Classificações por Tipo de Aluno



JNE/2013

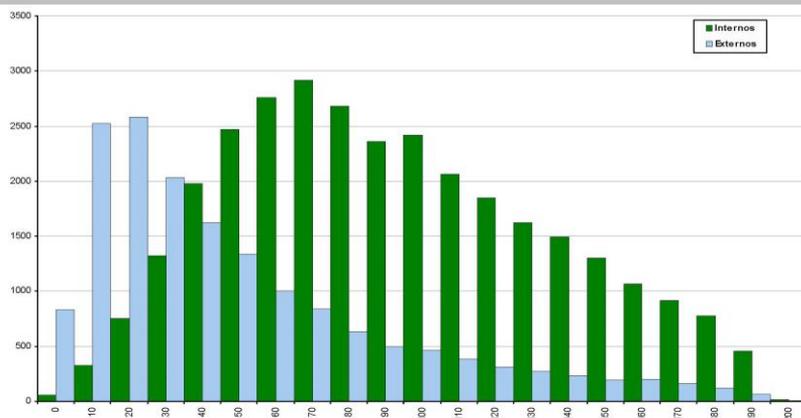
Página 5 de 5

Se observarmos a distribuição que se apresenta em seguida, as mesmas conclusões já não poderão ser tiradas apenas pela análise dos valores das médias. Ao levarmos apenas em linha de conta o valor de 82 pontos da média do exame de Matemática A (635) estamos a escamotear informações muito importantes relativamente ao desempenho dos alunos, nomeadamente, que existem muitos casos extremos, ou seja, muitos alunos com classificações muito elevadas e também com classificações muito baixas. Observa-se também uma grande diferença entre os alunos internos e os autopropostos. Esta distribuição só fica devidamente caracterizada através da análise fina da distribuição e das respetivas medidas de dispersão.

EXAMES FINAIS NACIONAIS DO ENSINO SECUNDÁRIO 2013

Distribuições de Classificações por Tipo de Aluno

635 Matemática A 1ª Fase



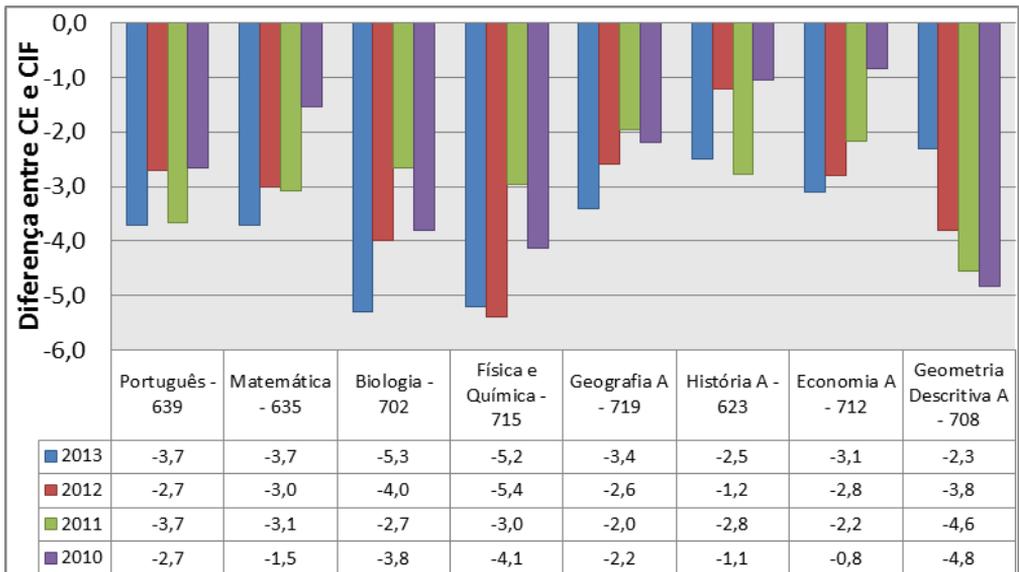
Ano / Fase	Inscrições	Provas	Falhas	Média	Internos	Média	Externos	Média
2013 / 1ª Fase	51951	47950	4001	82	31624	97	18328	64
2012 / 1ª Fase	54876	49831	4845	87	31450	104	15381	66
2011 / 1ª Fase	48952	39761	9191	92	28181	106	11580	67

JNE/2013

Página 16 de 18

Como já foi referido atrás a análise da diferença entre CIF e CE é de extrema importância para uma compreensão efetiva, tanto do desempenho de cada aluno, bem como de cada disciplina da escola, no que diz respeito ao nível micro, mas é também de enorme importância no sentido de determinar as desigualdades regionais e, principalmente as desigualdades entre tipos de escolas.

É considerado normal que exista uma diferença entre a avaliação interna e a externa, já que a CIF engloba outras aprendizagens não passíveis de serem testadas em exames escritos e corresponde à avaliação contínua dos dois ou três anos em que a disciplina é lecionada. Contudo, a diferença entre CIF e CE não pode ser demasiadamente acentuada pois que pode configurar problemas no processo de ensino e aprendizagem de uma disciplina, classificações internas demasiadamente inflacionadas ou, eventualmente, problemas nos instrumentos de avaliação externa.



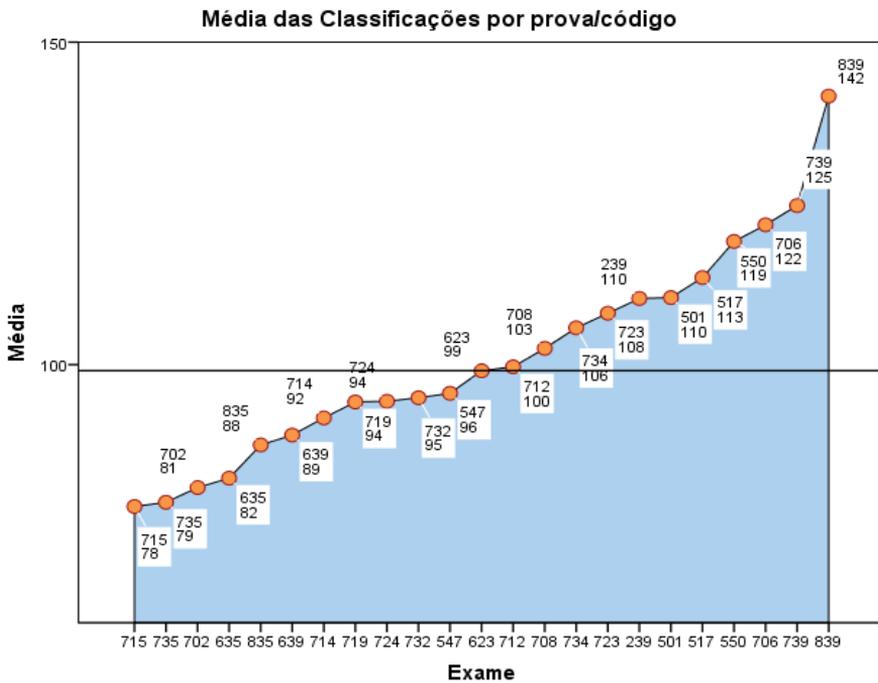
Como se pode observar no quadro, em algumas disciplinas, nomeadamente, Física e Química A (715) e Biologia e Geologia (702), esta diferença é muito pronunciada o que nos deve fazer refletir na procura de explicações e de soluções para este problema sempre latente e de grande impacto nos alunos, nas escolas e nos professores.

III – A avaliação externa regula os processos pedagógicos?

Observando o gráfico seguinte, podemos observar que algumas disciplinas têm médias das classificações de exame muito baixas, em comparação com as restantes. De entre estas podemos destacar as disciplinas de Física e Química A (715), Biologia e Geologia (702), Matemática A (635) e Português (639). Este panorama é intrigante por ser sistemático e consistente ao longo dos anos.

Será que são disciplinas mais difíceis? Será que o ensino nestas disciplinas é de má qualidade? Os professores que lecionam estas disciplinas não são competentes? Será que as práticas pedagógicas e as metodologias não são as adequadas? Os alunos têm, todos, dificuldades de aprendizagem nesta

área? Não parecem ser hipóteses explicativas viáveis. Tanto mais que se verifica nas escolas um enorme investimento em tempo e recursos humanos no ensino destas áreas.



Por outro lado, após uma análise rápida podemos observar que estas são as disciplinas requeridas para o concurso nacional de acesso ao ensino superior para os cursos de medicina, ou seja, são exames com um grau de seletividade extremamente elevado, tendo em consideração que o acesso a estes cursos se pode decidir por décimas. Esta constatação leva-nos a inferir que é provável que na elaboração destes exames haja a tendência natural para inserir mais itens de grau de dificuldade acima da média ou, então, que os critérios de classificação sejam mais exigentes que nos demais exames.

Tendo isto em consideração, e constatando a cada vez maior preocupação de professores e alunos com os maus resultados obtidos nestes exames a

nível nacional e o seu grau percebido de dificuldade, podemos assistir, principalmente nas ciências, a uma desvalorização de metodologias essenciais para a boa compreensão dos conceitos científicos, nomeadamente, o ensino experimental das ciências. Esta situação poderá dever-se ao facto de, à primeira vista, estas metodologias não serem diretamente valorizadas nos testes escritos, o que, não sendo necessariamente verdade, leva os professores a querer “preparar” os alunos para os exames sem “perder tempo” com atividades experimentais, dedicando todo o seu tempo ao treino puro e simples de problemas e exercícios.

Este é um bom exemplo de que nem sempre a avaliação externa contribui para uma melhoria das práticas pedagógicas e que deverá fazer refletir os decisores, os professores e as equipas que elaboram as provas.

IV – A avaliação externa contribui para intervenções diferenciadas no currículo?

A avaliação externa tem também como grande objetivo fornecer informação para que os decisores políticos possam elaborar ou ajustar as medidas de política educativa, necessárias para uma melhoria do sistema de ensino. Uma análise adequada e devidamente desagregada dos dados da avaliação externa poderá permitir encontrar padrões estatísticos que nos permitam efetuar intervenções no currículo destinadas à resolução de problemas específicos de grupos de alunos diferenciados. Como exemplos deste tipo de trabalho será interessante referir os seguintes grupos:

- Alunos do 1.º ciclo – Acompanhamento extraordinário do 1.º ciclo
- Cursos profissionais – prosseguimento de estudos no ensino superior
- Cursos científico-humanísticos – adequação do currículo aos vários cursos
- Alunos internos e autopropostos
- Estudos de género

Alunos do 1.º ciclo – Acompanhamento extraordinário do 1.º ciclo

A medida conhecida como Acompanhamento Extraordinário foi criada para apoiar os alunos que mostraram maiores dificuldades durante o ano escolar e que não conseguiram obter classificação igual ou superior a nível 3 a Português e ou Matemática. Fazendo uma análise dos dados da avaliação externa referentes a estes alunos, podemos obter informações muito importantes para que os decisores possam produzir eventuais alterações e ajustamentos.

No ano de 2013, no 4.º ano de escolaridade, encontravam-se em situação de não aprovação no ciclo 3856 alunos. Após a realização do Acompanhamento Extraordinário e a 2.ª fase das provas finais de ciclo recuperaram 652 alunos. Muitos consideram um número insuficiente, tendo em conta os recursos humanos envolvidos. No entanto, é preferível pensar neste número de forma absoluta, ou seja, 652 alunos conseguiram ter sucesso nas provas e transitaram para o 5.º ano de escolaridade. Parece-me que, só por isto, é uma medida com sucesso.

Por outro lado, esta medida despertou um enorme interesse da parte dos encarregados de educação pois que 67% dos alunos a Português e 77% dos alunos a Matemática que já se encontravam em situação de aprovação no ciclo mas com nível inferior a 3 a uma daquelas disciplinas, frequentaram o Acompanhamento Extraordinário e realizaram a prova na 2.ª fase para tentar recuperar. Em termos gerais, 10% dos alunos conseguiram obter classificação igual ou superior a nível 3 a Português na 2.ª fase. No caso da Matemática este número foi de 22% o que seria de esperar tendo em conta a duração desta medida e a didática específica desta disciplina.

Cursos profissionais – prosseguimento de estudos no ensino superior

As condições de acesso ao ensino superior dos alunos dos cursos profissionais sofreram algumas alterações no ano transato, passando estes a ter de realizar mais exames nacionais para cálculo da classificação final de curso para efeitos de prosseguimento de estudos. A maioria dos exames nacionais não correspondem às disciplinas que estes alunos têm no seu currículo e mesmo quando isso acontece, por exemplo o programa de

Português é idêntico, os conteúdos não são abordados da mesma forma. Uma análise comparativa dos dados da avaliação externa dos alunos dos cursos profissionais permite-nos verificar um desequilíbrio nas aprendizagens destes alunos em relação aos alunos dos cursos científico-humanísticos. Estes dados podem ser muito importantes para que possa proceder aos ajustamentos necessários nos currículos e também nas condições de acesso ao ensino superior para estes alunos.

Como se pode verificar no quadro seguinte, os alunos dos cursos profissionais não se encontram devidamente preparados para realizar, por exemplo, o exame nacional de Matemática A (635), já que obtiveram, na prova, uma média de 37 pontos, em 200, e, o que é mais marcante, uma mediana de 26 pontos, o que significa que metade dos alunos obteve classificação abaixo deste valor.

É esclarecedor também analisar o comportamento destes alunos no exame nacional de Português (639), dado que é um dos exames obrigatórios para prosseguimento de estudos. Apesar de o programa da disciplina ser igual, a média obtida pelos alunos dos cursos profissionais foi de apenas 63 pontos, contrastando com os 95 pontos dos alunos dos cursos científico-humanísticos, o que denota no mínimo a necessidade de alterações no processo de ensino e aprendizagem desta disciplina, nos cursos profissionais. O mesmo se pode dizer relativamente a Matemática B (735), cuja diferença de médias é substancial.

Prova/código	Tipos de curso	N	Média	Mediana
Matemática A (635)	Cursos científico-humanísticos	70609	86	80
	Cursos do ensino recorrente	2191	39	31
	Cursos do ensino profissional e tecnológico	2009	37	26
Português (639)	Cursos científico-humanísticos	76214	95	95
	Cursos do ensino artístico especializado	1062	81	78
	Cursos do ensino recorrente	2552	69	66

	Cursos do ensino profissional e tecnológico	14381	63	60
Matemática B (735)	Cursos científico-humanísticos	3600	84	80
	Cursos do ensino profissional e tecnológico	2496	57	50

Cursos científico-humanísticos – adequação do currículo aos vários cursos

Uma análise comparativa dos dados referentes aos quatro cursos científico-humanísticos permite-nos verificar diferenças que nos poderão ser úteis na adequação da aplicação dos programas das disciplinas aos currículos dos cursos e às características dos alunos, através de práticas pedagógicas diferenciadas. É de salientar que os alunos do curso de ciências e tecnologias têm melhores desempenhos a todas as disciplinas.

Prova/código	Curso	N	Média	Mediana
Matemática A (635)	C60 - Ciências e Tecnologias	60642	87	82
	C61 - Ciências Socioeconómicas	9446	77	71
	C62 - Línguas e Humanidades	89	61	45
	C64 - Artes Visuais	88	80	85
Português (639)	C60 - Ciências e Tecnologias	41774	101	100
	C61 - Ciências Socioeconómicas	6547	93	91
	C62 - Línguas e Humanidades	21216	88	86
	C64 - Artes Visuais	6360	81	77
Geometria Descritiva A (708)	C60 - Ciências e Tecnologias	3221	127	139
	C61 - Ciências Socioeconómicas	8	82	63
	C62 - Línguas e Humanidades	20	79	48
	C64 - Artes Visuais	7845	80	65
Filosofia (714)	C60 - Ciências e Tecnologias	4428	95	95
	C61 - Ciências Socioeconómicas	565	86	75

	C62 - Línguas e Humanidades	4033	89	84
	C64 - Artes Visuais	815	64	56
Geografia A (719)	C60 - Ciências e Tecnologias	328	97	98
	C61 - Ciências Socioeconómicas	6149	108	107
	C62 - Línguas e Humanidades	16369	92	90
	C64 - Artes Visuais	19	88	85

Alunos internos e autopropostos

Para uma análise adequada dos resultados dos exames nacionais é muito importante efetuarmos uma desagregação dos dados entre alunos internos e alunos autopropostos, tendo em consideração a significativa diferença entre os resultados destes dois grupos nas disciplinas com maior número de alunos.

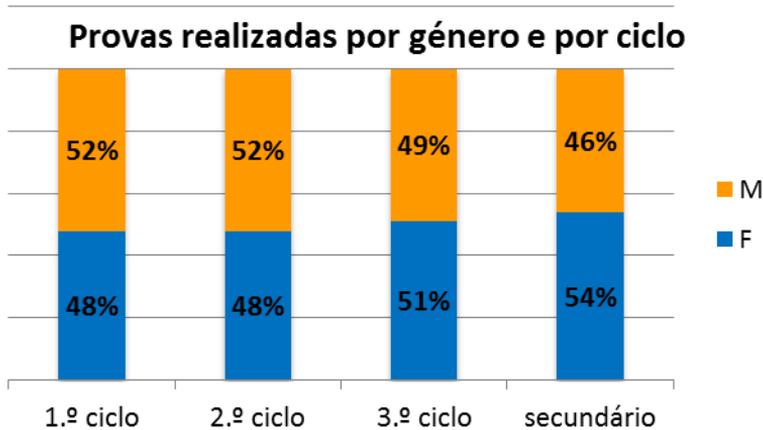
No quadro seguinte podemos ver que as diferenças entre internos e autopropostos em Matemática A é bastante pronunciada. No entanto, nas disciplinas de Biologia e Geologia e Física e Química A essa diferença já não é tão acentuada, pois que muitos alunos autopropostos realizam prova para melhoria de classificação nestas duas disciplinas bienais, respetivamente 22% e 18%, enquanto na disciplina de Matemática apenas 6% dos alunos realizam prova para melhoria de classificação. Assim, também é muito importante fazer a desagregação entre alunos autopropostos que realizam os exames para conclusão dos que os realizam para melhoria de classificação.

1.ª Fase		Autopropostos			Internos		
código	Prova	N	Média	Mediana	N	Média	Mediana
623	História A	4271	80	77	11546	106	105
635	Matemática A	16326	54	40	31624	97	95
639	Português	20723	67	65	50084	98	96
702	Biologia e Geologia	21555	76	70	29768	84	80

708	Geometria Descritiva A	3635	72	53	5648	122	127
712	Economia A	6032	88	85	5118	114	113
714	Filosofia	3297	75	65	5542	102	100
715	Física e Química A	22399	74	62	30499	81	75
719	Geografia A	4741	82	80	15145	98	97
735	Matemática B	3001	66	59	1674	102	100
835	MACS	2783	59	52	6628	99	99

Estudos de género

Os estudos de género são também muito relevantes, dado que os comportamentos e os desempenhos são muito diferenciados entre rapazes e raparigas. Se analisarmos o número de provas realizadas por género podemos obter informações muito valiosas para entender a influência das questões curriculares e didáticas em rapazes e raparigas. Como podemos observar no quadro seguinte o número de provas realizadas por rapazes no 1.º ciclo é maior do que o realizado por raparigas, o que é consistente com os dados estatísticos da natalidade. No 2.º ciclo ainda se verifica uma preponderância do sexo masculino relativamente ao feminino no que diz respeito ao número de provas realizadas. No entanto, no 3.º ciclo, verifica-se uma inversão no número de provas realizadas por género. Observa-se um maior número de raparigas a realizarem provas finais de ciclo o que pode indiciar uma maior dificuldade de os rapazes atingirem o 9.º ano de escolaridade ou então uma maior preponderância de alunos do género masculino em vias formativas alternativas, tais como os PCA, cursos CEF, cursos vocacionais ou PIEF, ou mesmo em abandono escolar. No ensino secundário esta diferença acentua-se, chegando a 54% do total. Esta situação é consistente com o número de alunos que acedem e terminam o ensino superior por género.



V – A avaliação externa promove a autoavaliação nas escolas?

Como já foi referido na primeira parte da comunicação, as escolas, para poderem realizar um processo de autoavaliação que possa ser operacional e efetivo, devem efetuar análises nos três níveis, macro, intermédio e local.

A nível local é fundamental analisar não só a avaliação obtida nas provas e exames realizados pelos alunos da escola mas também, e fundamentalmente, o ponto de partida de cada aluno, pelo que se torna absolutamente necessário que as escolas procedam a avaliação diagnóstica sistemática no início de cada ciclo. Assim, para poderem efetuar a sua autoavaliação as escolas deverão ter em consideração o seguinte:

- Estudar pormenorizadamente a distribuição de classificações nas várias disciplinas
- Comparar as CIF com as CE
- Comparar com os resultados dos anos anteriores
- Articular esta informação com estudos longitudinais relativos ao percurso escolar de cada aluno tendo em conta o seu ponto de partida em cada ciclo
- Ponderar com o contexto socioeconómico de cada aluno

VI – Os *rankings* de escolas dão-nos uma perceção real da qualidade das escolas?

A disponibilização dos dados da avaliação externa aos meios de comunicação social e ao público em geral contribui sobremaneira para transparência da ação do estado no que diz respeito ao sistema educativo, bem como para a prestação de contas aos cidadãos relativamente ao trabalho efetuado nas escolas.

No entanto, as análises simplistas baseadas apenas em comparação de médias das classificações de exame tornaram os *rankings* numa fonte de grande injustiça relativa entre as escolas, visto que uma escola que se encontre num meio socioeconómico menos favorecido tem maior probabilidade de os seus alunos terem um menor desempenho na realização das provas de avaliação externa. Este facto, porém, não significa que o trabalho realizado nessas escolas seja de menor qualidade. Por vezes, são escolas que se encontram inseridas em contextos sociais tão deprimidos que o seu trabalho primordial se localiza ao nível do combate ao abandono escolar dos seus alunos. Nestas escolas conseguir que um aluno termine o ensino secundário poderá ser considerado uma vitória.

Deste modo, deve ser de rejeitar de todo qualquer relação causa-efeito direta entre a média das classificações de exame e qualidade das escolas, já que não são tidos em conta outros fatores estatísticos importantes, nem os contextos socioculturais.

Apesar de tudo, atualmente encontramos-nos já algo longe dos *rankings* dos primeiros anos. Desde 2012 que o MEC começou a disponibilizar dados de contexto das escolas, a fim de que os meios de comunicação social possam entrar em linha de conta, na sua análise, com dados importantes tais como o contexto socioeconómico das escolas, o número de alunos da escola, o número de alunos por cada disciplina, entre outros. Os dados de contexto das escolas fornecidos aos *media* passam pela disponibilização de alguns indicadores, tais como:

- Percentagem de alunos que beneficiam de ASE
- Média do número de anos da habilitação dos pais

- Média do número de anos da habilitação das mães
- Percentagem de docentes do quadro de escola ou do quadro de zona pedagógica
- Melhoria estatisticamente significativa, com lugar à atribuição de crédito horário

VII – Conclusão

É um facto sobejamente aceite que a avaliação externa tem uma importância fulcral na regulação do sistema educativo. No entanto, o seu impacto não pode ficar restrito às suas funções usuais de certificação, aferição e seleção. Cada vez mais se torna fundamental saber utilizar e analisar os dados da avaliação externa para obter informação credível que nos permita melhorar o currículo, a avaliação e as didáticas, tendo em consideração todas as especificidades e diversidades que caracterizam o nosso sistema educativo.

MONITORIZAÇÃO E *BENCHMARKING* DE RESULTADOS ESCOLARES (EXAMES E PROVAS DE AFERIÇÃO) NA REDE ESCXEL

Rui Santos¹

A Rede de Escolas de Excelência (ESCXEL) é uma parceria iniciada em 2008, entre a Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa, através do seu Centro de Estudos de Sociologia (CESNOVA), e um conjunto de cinco municípios e respetivas escolas públicas. Hoje inclui oito municípios: Amadora, Batalha, Castelo Branco, Constância, Loulé, Mação, Oeiras e Vila de Rei, contando com a adesão voluntária de todos os agrupamentos de escolas e escolas não agrupadas existentes nos concelhos.

Somos uma rede colaborativa multifuncional, com uma grande diversidade de atividades que não cabe aqui resumir. No âmbito deste seminário, ater-me-ei ao processamento pelo CESNOVA dos resultados das avaliações externas dos alunos, expressas nas classificações médias nos exames nacionais de todos os ciclos do Ensino Básico e Secundário e, até recentemente, nas das provas de aferição no termo do 1.º e do 2.º ciclos do Ensino Básico (classificações externas adiante designadas genericamente por CE). Fazemo-lo através da construção de indicadores homogéneos que permitem:

- **Padronizar**, criando comparabilidade intertemporal.
- **Resumir** e simplificar padrões e tendências.
- **Mapear** desempenhos e, fundamentalmente, tendências de evolução.

O objetivo destes procedimentos e da disseminação dos indicadores processados pelos parceiros é múltiplo. Desde logo, o de acrescentar uma nova camada de inteligibilidade aos resultados, e especialmente à sua evolução, para informar a perceção, a problematização, a discussão e as

¹ CESNOVA, FCSH/NOVA

orientações de ação de professores e diretores, conselhos de escola, e responsáveis políticos e técnicos municipais.

Em segundo lugar, contribuir para uma cultura de colaboração e *benchmarking* entre os atores da rede, através da identificação de agrupamentos de escolas e das escolas não agrupadas (adiante designados genericamente por “escolas”, por comodidade de expressão) que apresentem tendências mais positivas para um dado nível médio de desempenho, apelando à identificação e à disseminação de boas práticas.

Finalmente, contribuir para a cultura de autoavaliação e para a apresentação de resultados às avaliações externas, sejam elas formais por parte da tutela, ou informais por parte da comunidade que servem.

Importa precisar o nosso papel neste processo: a equipa do CESNOVA tem a seu cargo a elaboração dos indicadores, a formação dos utilizadores para a sua leitura e a elaboração dos relatórios que os apresentam e os interpretam resumidamente. Cabe aos atores educativos e, particularmente, às escolas analisá-los e identificar os contextos e as práticas que lhes subjazem, discutir e partilhar experiências, para gizarem estratégias tendentes à melhoria contínua, que é no fundo o conceito de excelência que designa a rede.

Passo a apresentar em linhas gerais os principais indicadores que construímos, definindo o seu contributo para o acréscimo de inteligibilidade pretendido, segundo as três funções acima definidas.

Padronizar

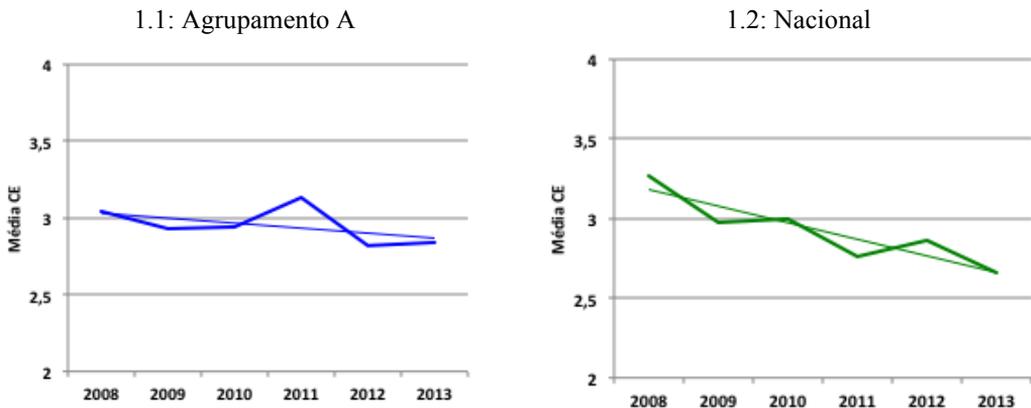
O primeiro princípio em que baseamos a construção dos nossos indicadores é a padronização dos resultados, de modo a que sejam comparáveis entre escolas e municípios e através do tempo, pois só assim podemos inferir conclusões seguras sobre os seus níveis e tendências de evolução.

De facto, cremos que a variabilidade interanual das médias nacionais das CE evidencia um efeito não negligenciável das variações das próprias provas, que pode viciar tanto a apreciação dos níveis médios de resultados

nas escolas e municípios num dado ano, como, e sobretudo, a da sua evolução.

Vejamos por exemplo o comportamento recente da média das CE do Agrupamento A da rede ESCXEL nos exames de Português no 3.º ciclo (fig. 1.1).

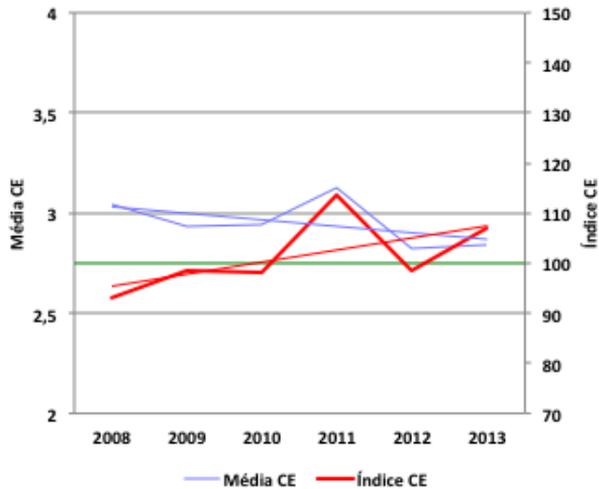
Figura 1: Médias das CE do Agrupamento A e nacional em Português no 9.º ano, 2008-2013



Fonte: Júri Nacional de Exames e ESCXEL, *Agrupamento [A] 2008-2013* (Lisboa: CESNOVA/FCSH, 2014).

A evolução das médias de resultados de exame no agrupamento configura uma evolução negativa ao longo dos seis anos. Contudo, essa evolução ganha um significado claramente diferente uma vez que a confrontamos com a da média nacional da CE, ainda mais negativa (fig. 1.2). Salvo admitindo que as competências dos jovens de todo o país em Português, à saída do 9.º ano de escolaridade, tenham simultaneamente oscilado e decrescido ao longo dos seis anos considerados, a evolução explica-se pela variação das provas de exame. Essa é a hipótese que assumimos. Assim, a evolução do Agrupamento A surge-nos completamente diferente, uma vez que os indexamos à CE média nacional (Fig. 2):

Figura 2: Índice da média da CE do Agrupamento A em Português no 9.º ano, base: média nacional da CE=100, 2008-2013



Fonte: Júri Nacional de Exames e ESCXEL, *Agrupamento [A] 2008-2013* (Lisboa: CESNOVA/FCSH, 2014).

Por este motivo, a maioria dos nossos indicadores de resultados escolares baseia-se em índices ou em rácios, tendo como base tornada constante a média nacional no ano de referência (valor 100 do índice CE, na figura 2). Refira-se que, por se tratar de uma indexação à média nacional que anula as diferenças interanuais devidas ao tipo de provas, este tratamento permitiu compatibilizar as novas séries de resultados dos exames nacionais do 4.º e do 6.º anos do Ensino Básico com as anteriores séries das provas de aferição.

Resumir

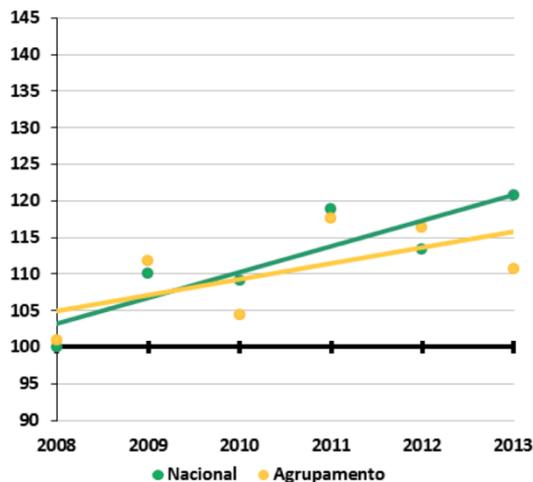
Construídos os indicadores, o seu tratamento visa resumir e simplificar a informação em tendências ao longo dos períodos considerados. Nos vários casos em que foram compostos ou recompostos agrupamentos de escolas ao longo de cada período em análise, agregámos retrospectivamente os resultados escolares das unidades (re)agrupadas para toda a janela de seis anos, segundo a composição do agrupamento no último ano. Tal como no exemplo acima, em cada relatório anual trabalhamos um período de seis

anos, funcionando numa lógica de janela móvel. Esse intervalo de tempo contém um número de pontos de dados suficiente para identificar uma tendência de evolução, e não tão amplo que essa tendência seja irrelevante para a avaliação das práticas atuais da escola. Esse resumo de tendências é feito à escala da escola e à escala do concelho, quando o concelho tem mais do que um agrupamento.

Relação entre avaliação externa e avaliação interna

Para medir a relação da classificação interna (CIF) com a CE e visualizar a sua evolução, construímos o índice CIF, em que a base 100 constante no tempo é a média CE, à qual indexámos a CIF. Nesta abordagem não assumimos a média nacional como constante, antes comparamos a evolução deste índice em cada disciplina entre a escola e o contexto nacional.

Figura 3: Valores e retas de regressão dos índices CIF do Agrupamento A e nacional em Português no 9.º ano, base: médias da CE=100, 2008-2013



Fonte: Júri Nacional de Exames e ESCXEL, *Agrupamento [A] 2008-2013* (Lisboa: CESNOVA/FCSH, 2014).

A figura 3 mostra-nos que, no Agrupamento A, a CIF em Português no 9.º ano esteve sempre acima da CE da disciplina, evoluindo no sentido da

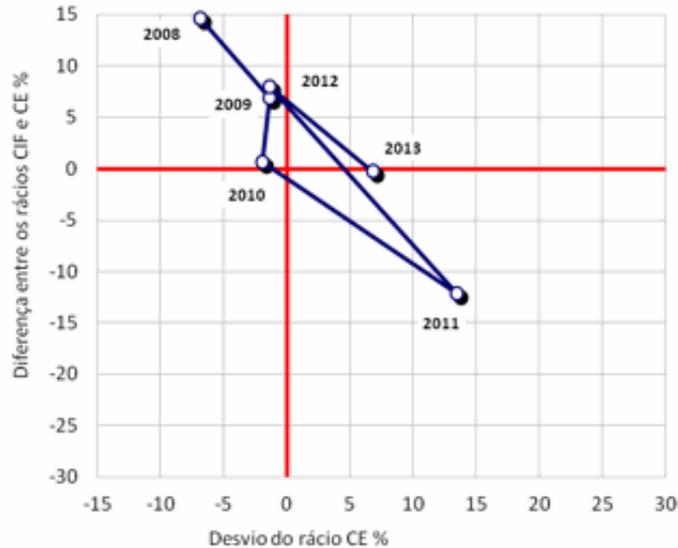
valorização relativa da CIF à medida que, como sabemos, as médias da CE desceram (fig. 1.1). À escala nacional, a valorização relativa da CIF foi em geral idêntica, mas a sua subida foi mais acentuada, à medida que as CE desceram também mais acentuadamente, como vimos (fig. 1.2), no fundo sugerindo a inelasticidade da CIF relativamente à CE em ambas as escalas, e maior à escala nacional. Isto mostra às escolas, não só o grau de ajustamento das suas práticas de avaliação interna relativamente aos critérios dos exames nacionais, como em que medida as suas práticas de avaliação interna a aproximam ou diferenciam da tendência nacional de relação entre estas duas classificações.

Numa abordagem diversa, resumimos a evolução da relação entre a CE e a CIF operando com dois indicadores de base: o rácio da média CE da unidade de análise (escola ou concelho) à média CE nacional (rácio CE) e o rácio homólogo da média CIF da unidade de análise à média nacional (rácio CIF).

A partir destes, construímos dois indicadores de diferença: o desvio do rácio CE relativamente ao valor de base nacional, que mede em que proporção a média CE da unidade de análise foi superior ou inferior à nacional; e a diferença entre os rácios CIF e CE, que mede a sub- ou sobrevalorização da CIF relativamente à CE, aferidas pela tendência nacional. A trajetória conjunta destes indicadores num diagrama de fases permite-nos visualizar em que medida a relação entre a sub- ou sobrevalorização da CIF e a CE, aferidas pelos valores nacionais, tendeu, relativamente à relação nacional:

- a permanecer:
 - num espaço de paridade;
 - num espaço de maior exigência da avaliação interna relativamente à externa;
 - num espaço de menor exigência da avaliação interna relativamente à externa;
- a evoluir num sentido ou no outro;
- a ter um comportamento errático.

Figura 4: Relação entre a diferença do rácio CE ao nacional e a diferença entre os rácios CIF e CE do Agrupamento A em Português no 9.º ano, 2008-2013



Fonte: Júri Nacional de Exames e ESCXEL, *Agrupamento [A] 2008-2013* (Lisboa: CESNOVA/FCSH, 2014).

O exemplo na fig. 4 mostra que no Agrupamento A, na disciplina de Português no 9.º ano do Ensino Básico, a relação entre a valorização relativa da CIF e a CE evoluiu entre 2008 e 2011 num sentido de aumento da exigência da CIF, por comparação com o ocorrido à escala nacional, passando ao longo destes quatro anos de uma posição de sobrevalorização relativa da CIF em 2008 para outro de clara sobrevalorização da CE. Nos dois últimos anos do sexénio, regressou a posições mais próximas da relação observada à escala nacional, no triângulo delimitado pelos pontos de 2009, 2010, 2012 e 2013 e para o qual tornou sempre após as posições extremas de 2008 e 2011: valorização relativa da CIF moderada ou nula, combinada com médias CE quase coincidentes, ou moderadamente abaixo das nacionais. A escola pode assim confrontar esta informação com as suas estratégias de avaliação interna na disciplina, e aferir se esta evolução corresponde aos seus objetivos.

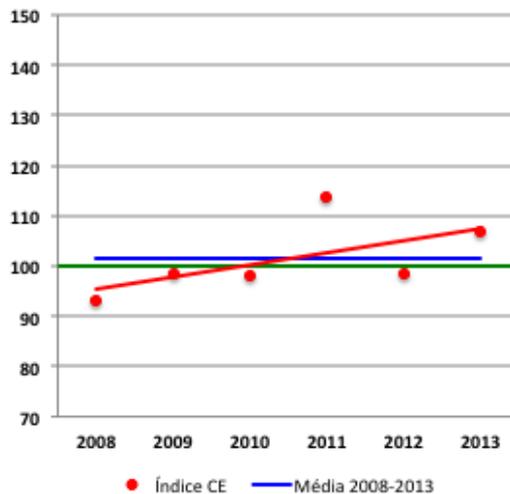
Nível relativo e evolução dos resultados da avaliação externa

Para resumir a evolução dos resultados da avaliação externa, usamos dois indicadores principais:

- o nível médio dos resultados da unidade de análise na avaliação externa, na média das disciplinas e em cada disciplina, relativamente aos resultados nacionais (ou seja, a média do índice CE nos seis anos em análise);
- a progressão do nível de resultados ao longo do período, dado pelo pendor, positivo ou negativo, e pelo declive mais ou menos acentuado da reta de regressão ajustada aos valores do índice CE ao longo dos anos em análise.

A figura 5 visualiza esses dois indicadores, mostrando que o Agrupamento A teve, no conjunto do período, um nível médio ligeiramente superior ao nacional nos resultados da CE de Língua Portuguesa de 9.º ano, e evidenciou ainda uma evolução ascendente, mercê do baixo valor de partida e dos melhores resultados em 2011 e 2013.

Figura 5: Valores anuais, média e reta de regressão do índice CE do Agrupamento A em Português no 9.º ano, 2008-2013



Fonte: Júri Nacional de Exames e ESCXEL, *Relatório do Agrupamento [A] 2008-2013* (Lisboa: CESNOVA/FCSH, 2014).

Mapear desempenhos e tendências de evolução

Finalmente, para mapear de modo sintético os desempenhos e as tendências de evolução, utilizamos *scoreboards* (tabelas de pontuações) em que a tendência é definida pela combinação da média dos índices CE do período considerado com o declive da reta de regressão que resume a sua progressão, segundo as seguintes escalas (fig. 6 e 7):

- atribuímos o *score* 1, “bom”, quando a média dos índices CE nos anos considerados é igual ou superior à média nacional da prova (valor 100);
- atribuímos o *score* 2, “sofrível”, quando a média é inferior em até 5%, inclusive;
- atribuímos o *score* 3, “em risco”, quando a média é inferior em mais de 5% e menos de 15%;
- atribuímos o *score* 4, “mau”, quando a média é inferior em 15% ou mais.

Figura 6: *Scores* atribuídos à média, segundo os intervalos de médias dos índices CE

Média	Score
≥ 100	1
$[95,100[$	2
$]85,95[$	3
≤ 85	4

O declive, por sua vez, resume a progressão dos resultados, representando a variação linear de a pontos percentuais no índice por cada ano do período observado. No caso de um declive positivo (a superior a 0), a reta é ascendente, o que significa que, no conjunto dos dados observados, a progressão tendencial dos índices CE foi à razão de $a\%$ em cada ano. Inversamente, no caso de um declive negativo (a inferior a 0), a reta é descendente, o que significa que os índices CE tenderam a diminuir $a\%$ em cada ano. No caso de um declive nulo (a igual a 0), a reta é horizontal.

Isto significa que, atravessando as variações interanuais, o sentido da progressão dos resultados foi de estagnação ao longo dos anos em análise. Tal como fizemos para a média dos índices CE, atribuímos *scores* a intervalos de valores do declive (fig. 7):

- atribuímos o *score* 1, “bom”, quando o declive foi igual ou superior a 2, ou seja, uma progressão tendencial positiva de 2% ou mais por ano;
- atribuímos o *score* 2, “sofrível”, quando o declive do Agrupamento foi inferior a 2 e superior a 0, ou seja, uma progressão tendencial positiva, mas fraca;
- atribuímos o *score* 3, “em risco”, quando o declive do Agrupamento foi nulo ou negativo mas superior a -2, ou seja, uma progressão tendencial estagnada ou moderadamente negativa;
- estão classificados com o *score* 4, “mau”, os casos em que o declive do Agrupamento foi negativo e igual ou inferior a -2, ou seja, uma progressão francamente negativa, com decréscimo tendencial de 2% ou mais por ano.

•

Figura 7: *Scores* atribuídos ao declive, segundo os intervalos de declives da reta de regressão dos índices CE

Declive	Score
≥ 2	1
$]0,2[$	2
$] -2,0]$	3
≤ -2	4

A tendência, finalmente, resulta da combinação dos dois *scores* acima. Como mostra a figura 8, supondo que o ritmo de progressão se mantivesse no futuro próximo, consideramos que mesmo uma escola com resultados iguais ou superiores à média nacional estará em risco se tiver tido uma progressão muito negativa (*scores* [1,4]), e uma tendência sofrível se piorou um pouco ou não conseguiu melhorar (1,3); permanecendo com

uma boa tendência se tiver melhorado mesmo marginalmente ([1,1] ou [1,2]).

Reciprocamente, uma escola com resultados muito inferiores à média nacional terá, apesar de uma média má, uma tendência sofrível no curto prazo se conseguiu uma evolução francamente positiva (4,1), e menos má, embora ainda em risco, se tiver tido uma progressão positiva sofrível (4,2); mas já se manterá com uma tendência má se não tiver conseguido melhorar nada, ou tiver piorado mesmo que marginalmente ([4,3] ou [4,4]). *Mutatis mutandis*, a mesma lógica aplica-se às situações intermédias.

Figura 8: Scores atribuídos à tendência, segundo as combinações dos scores da média e do declive

Média	Declive	Tendência
1	1	1
1	2	1
2	1	1
1	3	2
2	2	2
3	1	2
4	1	2
1	4	3
2	3	3
3	2	3
4	2	3
2	4	4
3	3	4
3	4	4
4	3	4
4	4	4

A tabela apresenta os scores em sentido descendente, de 1 (verde), os melhores resultados, a 4 (vermelho), os piores. É com estes blocos de cor

que construímos os *scoreboards*, para duas escalas (por escola e por concelho), com diferentes agregações de resultados (por ciclo e por disciplinas em cada ciclo), e representados com diferentes graus de complexidade (desde a média dos índices de todos os exames de só um ciclo até todos os ciclos com as disciplinas discriminadas). A tradução em cores, conforme apresentado nas figuras antecedentes, tornam-nos instrumentos expressivos que permitem uma leitura gráfica rápida dos resultados médios, das suas evoluções e das tendências resultantes no conjunto da rede.

Na ordenação dos *scoreboards* consideramos como primeiro critério as tendências, que resumem as duas variáveis estatísticas de média e declive e, como segundo critério, a média, que corresponde à posição da escola em relação aos resultados médios nacionais dos exames. Esta opção visa acentuar a capacidade de melhorar os resultados, tendo embora em conta o nível médio destes que, como ficou dito, por sua vez resulta em parte da sua evolução.

Tratando-se de instrumentos de leitura simplificados, estes *scoreboards* não dispensam a leitura e a análise dos dados subjacentes; sob os mesmos *scores* de média e declive, podem encontrar-se distribuições substancialmente diferentes de resultados concretos. Por outro lado, pese embora a ordenação, não pretendemos que sejam utilizados como meros *rankings* de classificação de concelhos ou escolas. Trata-se, sim, de mapas comparativos que resumem o estado dos resultados da rede no cômputo geral do sexénio de referência, em que cada escola pode situar-se e, sobretudo, identificar outras escolas da rede com as quais se possa comparar mais diretamente, tendo em vista o objetivo mais importante: a utilização da rede para *benchmarking*, ou seja, para a aprendizagem partilhada de práticas que a experiência mostre terem resultado bem em situações comparáveis (ou, inversamente, de erros a evitar).

A figura 9 representa o mais recente *scoreboard* da rede relativo ao 9.º ano, por escolas, com o período de referência de 2008-2013. Num relance, podemos ver que no conjunto da rede, as escolas apresentam melhores tendências na disciplina de Matemática do que na de Português, e que isso

se deve não só a melhores médias (em 25 escolas, 11 tiveram médias boas em Matemática, e apenas nove em Português) mas também, e sobretudo, a melhores progressões dos resultados. Com efeito, seis escolas apresentam bons declives em Matemática, face a apenas uma em Português; disciplina esta em que houve igualmente um caso de declive francamente negativo, ausente na de Matemática. Não menos importante é o facto de só quatro das oito escolas com médias más ou em risco na disciplina de Português apresentarem declives positivos, ainda que sofríveis; enquanto em Matemática, de nove escolas com médias más ou em risco, seis tiveram progressões sofríveis.

Por outro lado, este *scoreboard* mostra que em ambas as disciplinas, as progressões positivas se concentraram nas escolas com médias positivas (12 de 16 em Português e 11 de 17 em Matemática). Em parte, isso deve-se à relação entre o declive e a média. Como ambos os indicadores foram computados para o mesmo intervalo de tempo, a média não é um valor de partida, mas de chegada. Para situações à partida idênticas, um declive positivo traduz-se num aumento da média, e um negativo numa redução. Logo, se o declive for acentuado, num sentido ou noutro, estará mais provavelmente associado a um melhor ou pior *score* da média. Ainda assim, não deixa de ser preocupante a escassez de declives sofríveis, e a ausência de bons, entre as escolas com médias em risco ou más. Sobretudo atendendo a que, pela relação entre o declive e a média, nas escolas com maus resultados à partida as progressões positivas se traduziriam mais provavelmente em médias de risco ou sofríveis; ora, mesmo entre as escolas com médias sofríveis, só uma (a escola M), e esta apenas em Matemática, apresenta um bom declive de progressão.

Figura 9: *Scoreboard* de Português no 9.º ano por agrupamentos, 2008-2013

Concelho	Agrupamento / Escola	PORTUGUÊS			MATEMÁTICA		
		M	D	T	M	D	T
A	A	Green	Green	Green	Green	Green	Green
B	B	Green	Light Green	Green	Green	Green	Green
A	C	Green	Light Green	Green	Green	Green	Green
A	D	Green	Light Green	Green	Green	Light Green	Green
B	E	Green	Light Green	Green	Green	Yellow	Light Green
A	F	Green	Yellow	Light Green	Green	Light Green	Green
A	G	Light Green	Light Green	Light Green	Green	Green	Green
C	H	Light Green	Light Green	Light Green	Green	Light Green	Green
C	I	Green	Yellow	Light Green	Green	Yellow	Light Green
B	J	Green	Yellow	Light Green	Green	Yellow	Light Green
D	K	Yellow	Light Green	Yellow	Green	Green	Green
E	L	Light Green					
F	M	Yellow	Light Green	Yellow	Light Green	Green	Green
A	N	Light Green	Light Green	Light Green	Light Green	Yellow	Yellow
D	O	Light Green	Yellow	Yellow	Light Green	Light Green	Light Green
D	P	Light Green	Light Green	Light Green	Yellow	Light Green	Yellow
G	Q	Light Green	Light Green	Light Green	Yellow	Light Green	Yellow
A	R	Light Green					
A	S	Green	Red	Yellow	Light Green	Yellow	Light Green
D	T	Yellow	Light Green	Yellow	Light Green	Light Green	Yellow
D	U	Yellow	Light Green	Yellow	Red	Light Green	Yellow
A	V	Yellow	Yellow	Red	Yellow	Light Green	Yellow
A	W	Yellow	Yellow	Red	Yellow	Yellow	Red
B	X	Yellow	Yellow	Red	Yellow	Yellow	Red
A	Y	Red	Yellow	Red	Red	Yellow	Red

Fonte: ESCXEL, Relatórios 2008-2013 (Lisboa: CESNOVA/FCSH, 2014).

Numa ótica de *benchmarking*, mais específica a cada escola, poderão por exemplo as escolas W, X e Y, com tendências más nas duas disciplinas, e a V com tendência má em Português, procurar junto das escolas T e U, com médias de risco ou más, identificar o que fizeram para terem alcançado declives positivos em ambas as disciplinas, assim como a K e a

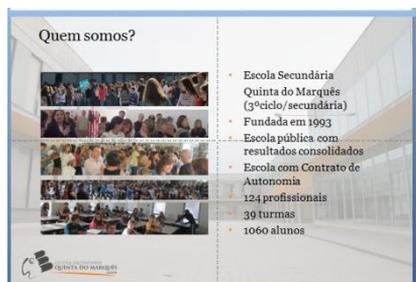
M em Português e as P, Q, R e V em Matemática, etc. O *scoreboard* pode assim servir como um mapa para cada escola identificar possíveis parcerias para aprendizagem dentro da rede, dentro e fora dos seminários que realizamos regularmente.

Observações finais

Os resultados de exames nacionais não são mais do que uma parte da avaliação externa do trabalho das escolas, por um lado, e da avaliação dos conhecimentos e das competências dos alunos, por outro. Num e noutro terreno, não devem ser tomados como dados absolutos e irredutíveis; as suas determinantes, por outro lado, são certamente complexas e insuficientemente conhecidas. Não podem, contudo, ser ignorados, antes devem ser trabalhados, aprofundada a sua análise e reforçada a sua problematização. O trabalho que prosseguimos na rede ESCXEL e que aqui apresentei em traços largos visa isso mesmo: apresentar de uma outra forma às escolas e aos municípios os dados que, à partida, todos conhecem, processados e organizados de modo a ir além do óbvio, a reduzir a complexidade e, acima de tudo, a estimular a comparação, a interrogação e a colaboração como ferramentas contra a complacência ou o conformismo.

Júlia Tainha¹

Quem somos?



A Escola Secundária Quinta do Marquês (ESQM), em Oeiras, é uma escola secundária com 3.º ciclo, uma escola pública com resultados consolidados, com Contrato de Autonomia desde 2007, uma escola com 124 profissionais, 39 turmas e 1060 alunos.

A centralidade do Projeto da escola reside no ALUNO: é para ele que se trabalha; é em função dele que a escola se tem vindo a adaptar constantemente. Nestes 20 anos de existência, o **projeto** da escola foi-se modificando, procurando **ajustar-se às mudanças**, quer do contexto externo, quer da população discente e malha urbana envolvente, e procurando corresponder aos desafios, mantendo, no entanto, as suas linhas orientadoras estruturantes.

A ESQM é uma escola secundária com 3.º ciclo. Enquanto escola pública, quer oferecer um serviço de qualidade, promover aprendizagens significativas, apostar no **desenvolvimento pessoal e social** e na **motivação para a aprendizagem**, numa fase fundamental de construção identitária dos jovens. Tem cumprido o que consideramos ser um **serviço público de qualidade**.

As opções da ESQM e o trabalho desenvolvido pelos seus professores e técnicos ao longo do tempo têm trazido resultados educativos bastante positivos, já consolidados, que se revelam não só ao nível das classificações finais das disciplinas, mas também na boa preparação dos

¹ Diretora da Escola Secundária da Quinta do Marquês, Oeiras

alunos para o ciclo de estudos posteriores, na sua formação integral e na sua satisfação e sentido de pertença à escola. Os seis anos que os alunos passam na escola criam um maior sentido de responsabilidade; a comunidade reconhece a escola e confia nos seus profissionais e no seu projeto.

A **introdução dos exames**, em 1997 no secundário e em 2005 no básico, obrigou a escola a centrar-se mais nas disciplinas específicas com piores resultados nos exames e a desenvolver estratégias para melhorar o desempenho dos alunos. O **impacto dos resultados** dos primeiros exames na escola foi enorme e **os resultados obtidos muito aquém do que esperávamos**. O desaire fez-nos refletir nas melhores estratégias para superar essas dificuldades.

Nos últimos 15 anos, a ESQM posiciona-se sempre entre as dez melhores escolas públicas do ensino secundário, muitas vezes entre as cinco melhores e, no ensino básico, nos últimos cinco anos, entre as cinco primeiras da Grande Lisboa e entre as melhores do país. Os resultados dos exames podem, de facto, ser um incentivo para melhorar a qualidade das aprendizagens, como a nossa escola o demonstra. É preciso que o aluno esteja no centro das preocupações da escola.

Missão

Tem-nos guiado a preocupação de alcançar **bons resultados** a nível educativo. A ESQM tem criado e recriado projetos para dar o seu contributo inequívoco para o **desenvolvimento pessoal, cultural e social** dos jovens que a frequentam. A ESQM assumiu-se como uma



escola verdadeiramente cultural, virada ao exterior, proporcionando aos alunos experiências de diversa índole: cultural, artística, cívica... Embora sem os recursos físicos adequados (escola com pouco espaço, sem ginásio, sem qualquer infraestrutura de carácter social para alunos ou professores), aproveitou todos os espaços a nível do currículo formal e não formal para concretizar esse desígnio. Aproveitou as virtualidades do Contrato de

Autonomia para obter instalações mais condignas e o tão necessário ginásio e espaço multiusos. No entanto, ao fim de 20 anos, continuamos sem estes espaços tão fundamentais. Continuamos a “reinventar” a escola, de forma a que esta motive para as aprendizagens.

Após a primeira avaliação externa das escolas em 2006, à qual nos candidatamos, o Contrato de Autonomia veio viabilizar, numa primeira fase, a estabilização de docentes na escola e a continuidade de projetos, a reabilitação da Matemática e de outras disciplinas e a melhoria substancial dos resultados.

A escola tem responsabilidade no **desenvolvimento do potencial intelectual, afetivo, social dos seus alunos**; à ESQM interessam os **resultados académicos**, mas sobretudo o desenvolvimento integral e harmonioso dos alunos, para que cada um desenvolva ao máximo as suas capacidades, como é **missão** da escola — melhorar o potencial da escola significa melhorar o potencial de cada indivíduo, de cada aluno.

Aos alunos procura-se inculcar a importância de aprender a aprender e da aprendizagem ao longo da vida; a ESQM tem-se ajustado, tem procurado soluções, a nível profissional e organizacional, que lhe permita cumprir eficientemente a sua **missão**, responder às necessidades de educação e formação dos seus alunos e às exigências da sociedade. São estas últimas que fundamentam a oferta educativa da escola, maioritariamente orientada para o **prosseguimento de estudos**. E é o projeto educativo da escola que justifica que esta oferta se não limite a cumprir os planos de estudo e os programas e a satisfazer-se apenas com os resultados dos exames nacionais e do concurso de acesso ao ensino superior e procure, quer no âmbito das atividades de complemento curricular, quer dos próprios planos de estudo e programas, promover uma **formação cultural ampla e diversificada**, articulada com a **formação cívica** que os jovens necessitam.

Visão



A ESQM criou salas de estudo, *ateliers* de tempos livres, espaços de apoio para os alunos, atividades e projetos — sempre procurou compatibilizar uma abordagem mais informal, que promovesse uma educação alargada, com a preparação dos alunos para obterem bons

resultados académicos.

A **escolaridade obrigatória de 12 anos** também leva a escola a repensar algumas práticas educativas. Todas as medidas integradoras, diferenciadas e criativas, concebidas e implementadas, dirigidas aos alunos abrangidos pela escolaridade obrigatória até ao momento presente, terão a sua continuidade, mas, a partir de agora, considerando não só o 3.º ciclo do ensino básico, mas também o ensino secundário. Acresce aqui uma mudança de paradigma e um enorme desafio para a escola, pois terá que se alcançar o **equilíbrio entre um ensino secundário de qualidade, que permita o prosseguimento de estudos, e a capacidade de desenvolver estratégias e dinâmicas que garantam a conclusão do ensino secundário para todos**. Se, até há pouco tempo, a conclusão do ensino secundário foi uma meta a que a escola sempre se propôs, hoje, é um pressuposto para todas as escolas do país, cabendo-nos a nós encontrar formas interessantes e criativas de alcançar esta meta com a qualidade necessária, para que o ensino secundário seja uma base sólida que possa alavancar percursos educativos e profissionais posteriores. É mais um desafio que aceitamos.

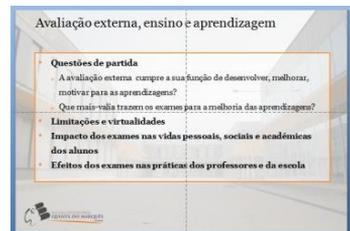
No entanto, sentimos cada vez mais constrangimentos à nossa ação; o que as escolas faziam já não conseguem fazer. Temos cada vez menos recursos e menos flexibilidade, menos capacidade ou margem de manobra. Poderíamos pensar que o Contrato de Autonomia proporciona esses espaços de autonomia, mas está muito aquém de corresponder às nossas expectativas e necessidades.

A ESQM preocupa-se, na verdade, em proporcionar aos seus alunos aprendizagens de qualidade, para que obtenham “*um elevado grau de realização escolar e de desenvolvimento pessoal*”.

Não se ensina só para os exames, na medida em que se procura uma formação integral, mas não se descarta o papel dos exames no ensino e na aprendizagem, não se descutando, igualmente, a sua importância no futuro dos alunos, nas condições de acesso ao ensino superior e à profissão desejada.

Avaliação Externa, Ensino e Aprendizagem

A **temática** deste Seminário “Avaliação Externa e Qualidade das Aprendizagens”, tendo por base a realidade **EXAMES**, remete-nos, de imediato, para a questão: será que a avaliação externa cumpre a sua função de **desenvolver, melhorar, motivar para as aprendizagens**? Que **mais-valia** trazem os exames para a melhoria das aprendizagens?



A preocupação da ESQM tem sido organizar-se da melhor forma para **responder às necessidades dos nossos alunos**. A ESQM é uma escola de pressão; confronta-se com o dobro das candidaturas nos anos de entrada. A população escolar é competitiva, maioritariamente virada para o prosseguimento de estudos. As expectativas das famílias são altas. A escola tem o dever de assegurar condições iguais para todos, pelo que procura incutir nos seus alunos uma cultura de trabalho e rigor, para fazer mais e melhor, e procura, igualmente, oferecer os recursos de apoio necessários aos que apresentem maiores dificuldades.

Na ESQM, conscientes da importância dos exames para a certificação dos jovens, para o mercado de trabalho e, sobretudo, para o prosseguimento de estudos, temos procurado preparar os alunos para fazer face aos exames e obterem bons resultados. Os Serviços de Psicologia e Orientação utilizam,

igualmente, esses resultados como um dado importante no aconselhamento e preparação do aluno para ingresso no ensino superior.

Naturalmente, há **limitações e virtualidades** em todo este processo. Os exames trazem limitações ao nível dos alunos, das práticas de ensino e da orgânica da escola. São reguladores do processo de ensino e aprendizagem; a ação do professor fica condicionada pelo exame, no tipo de tarefas que propõe aos alunos, nos conteúdos específicos a abordar, nas metodologias de sala de aula. Os professores têm desenvolvido competências para prepararem os alunos para um bom desempenho nos exames. Na ESQM, para além de outras medidas tomadas, aderimos também aos testes intermédios como uma boa estratégia de preparação para os exames, do seu ponto de vista formativo e sumativo.

Mais uma vez, cabe à escola encontrar o equilíbrio entre esta realidade e a necessidade de metodologias mais ativas e outras formas de avaliação formativa; a necessidade de articular a avaliação interna com a avaliação externa.

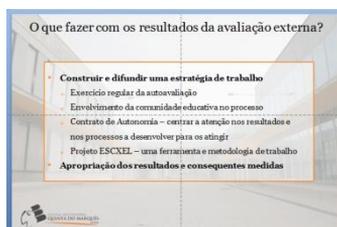
A introdução dos **Rankings** tem sido, igualmente, um elemento de pressão; podemos também refletir sobre o desempenho das escolas públicas face às privadas, sendo que os recursos e o público-alvo de umas e de outras são substancialmente diferentes, não comparáveis. Os exames não permitem igualmente uma comparação ano a ano; o seu grau de dificuldade varia, não estão devidamente calibrados e há muitos outros fatores que influenciam os resultados.

Na nossa escola procurámos utilizar internamente esses dados e os resultados disponibilizados e refletir sobre a posição relativa da escola e das diferentes disciplinas nesses *rankings*. Os *rankings*, complementados com os documentos elaborados na escola sobre os resultados da avaliação interna e externa, são motivo de reflexão do Conselho Pedagógico, aprofundada pelos departamentos, tomando por base, por um lado, os indicadores gerais e as linhas de análise estabelecidos e, por outro, os resultados pormenorizados de cada uma das turmas. É, assim, possível **identificar tendências, desvios e pontos críticos** que cada um dos

departamentos e grupos considere relevantes e propor as medidas mais ajustadas para a melhoria dos resultados e, conseqüentemente, mobilizar os recursos para melhorar o desempenho da escola.

O que fazer com os resultados da avaliação externa?

A ESQM desde sempre tomou como diretrizes o **exercício regular da autoavaliação**, com a sistemática identificação dos pontos fortes e dos pontos fracos; a procura de indicadores relevantes para análise dos resultados; apreciação e discussão dos resultados em Conselho Pedagógico, Departamento, Grupo disciplinar. Preocupação em devolver aos professores e profissionais da escola os resultados, pois são eles os maiores agentes de mudança. A autoavaliação tem sido uma prática regular.



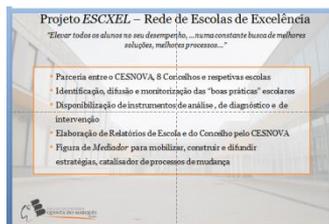
Em 2006 a Avaliação Externa, à qual nos propusemos, trouxe-nos outro incentivo para procurar mudanças concretas. Daí resultou um dos primeiros 22 Contratos de Autonomia do país. Um olhar exterior sobre os resultados, a pressão exercida pelo Contrato de Autonomia obrigaram a comunidade escolar interna a estar ainda mais atenta aos resultados dos alunos e ao seu próprio desempenho. O **Contrato de Autonomia** centrou a atenção nos resultados, mas também, naturalmente, nos processos a desenvolver para os atingir. Levou-nos a estabelecer objetivos mais concretos e quantificáveis. Os professores, no âmbito dos seus departamentos, analisavam os resultados, procurando interpretar os desvios e encetar estratégias ao nível do seu grupo/disciplina, para melhorar o desempenho dos alunos.

A apropriação dos resultados por parte dos intervenientes diretos conduziu, necessariamente, a novas responsabilidades e conseqüentes mudanças nas práticas pedagógicas e organizacionais da escola.

A escola integrou o Projeto ESCXEL em 2008, o que se revelou uma excelente ferramenta de trabalho.

Projeto ESCXEL – Rede de Escolas de Excelência

”Elevar todos os alunos no seu desempenho, ...numa constante busca de melhores soluções, melhores processos...”



A ESQM integra o “**Projeto Escxel**”, desde 2008, em parceria com o CESNOVA e as escolas de cinco concelhos envolvidos, prosseguindo os mesmos objetivos de melhoria da escola e do seu desempenho.

As diferentes vertentes do Projeto ESCXEL, identificação, difusão e monitorização das “boas práticas” escolares; disponibilização de instrumentos de análise de diagnóstico e de intervenção; elaboração de Relatórios de Escola e do Concelho pelo CESNOVA; figura de *Mediador* para mobilizar, construir e difundir estratégias, catalisador de processos de mudança, trouxe uma grande mais-valia à escola. O projeto tem sido propiciador de oportunidades de aprendizagem, momentos de reflexão e de trabalho cooperativo, com diversas valências e envolvimento dos elementos da comunidade educativa. É parte integrante do Projeto Educativo da escola, contribuindo para alcançar as metas que nos propomos.

A escola já estava habituada a olhar para os seus resultados, a envolver os professores na interpretação das tendências e desvios e na procura de novas estratégias para colmatar as dificuldades; os relatórios produzidos pelo CESNOVA sobre os resultados dos exames na ESQM reforçaram essa análise.

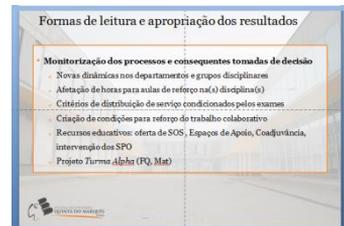
Como estratégia de trabalho criamos momentos de análise e reflexão por parte dos responsáveis pela gestão da escola; sessões/debate envolvendo todos os professores das diferentes disciplinas; recolha de elementos para a interpretação dos dados; identificação de áreas de fragilidade; sugestões de melhoria; apresentação da análise dos dados recolhidos.

A escola deu um **salto qualitativo na análise e interpretação dos resultados**. A introdução de outros indicadores e elementos de análise e de reflexão permitiu uma reflexão mais profunda, uma análise

comparativa com anos anteriores, permitiu evidenciar factos disciplina a disciplina, permitindo clarificar os objetivos e metas a alcançar. Deixamos de ver os resultados, as médias em abstrato; permitiu uma interpretação mais detalhada dos fatores que influenciaram os resultados disciplina a disciplina e em contexto; a comparação da nota interna e externa. Evidenciou as **margens de progresso** muito mais claramente.

Formas de leitura e apropriação dos resultados

Esta análise constante dos resultados e dos desvios destes relativamente aos resultados esperados e aos resultados pretendidos levamos a procurar implementar ações concretas e mobilizar todos os recursos ao nosso alcance para melhorar as nossas práticas.



Os resultados dos exames nacionais, quer no caso do secundário, na segunda metade dos anos 90, quer no caso do básico já mais recentemente, foram fracos, como já referido anteriormente, abaixo das expectativas. A reflexão que a direção alargada fez relativamente às causas da situação com que se viu defrontada em cada um desses dois momentos mobilizou-nos para **ações concretas**: a reorganização dos departamentos e a aposta no planeamento conjunto e parcerias; a continuidade pedagógica e a consequente responsabilização de cada um pelos resultados dos seus alunos, pelo menos num ciclo de estudos; a distribuição de professores pelo básico e secundário (a aposta da escola da mesma tipologia, 3.º ciclo e secundário, tem trazido aspetos muito positivos); trabalho intenso no PAM, Plano de Acompanhamento da Matemática e adesão ao Projeto-piloto dos Novos Programas da Matemática do Ensino Básico; aposta na ação estratégica da Biblioteca e desenvolvimento da leitura, antes e após o Plano Nacional da Leitura; candidatura a projetos Comenius e outros...

Em suma, como resultado desta avaliação constante dos resultados, a gestão tem mobilizado os recursos necessários, ou melhor, os recursos possíveis, para melhorar o desempenho da escola:

- Novas dinâmicas nos departamentos e grupos disciplinares;

- Afetação de horas para aulas de reforço na(s) disciplina(s);
- Critérios de distribuição de serviço condicionados pelos exames;
- Criação de condições para reforço do trabalho colaborativo;
- Recursos educativos: oferta de SOS, Espaços de Apoio, Coadjuvância, intervenção dos SPO;
- Projeto *Turma Alpha* (FQ, Mat).

Projeto Turma Alpha – Como funciona?



Face à leitura dos resultados, a escola mobiliza-se para achar as melhores soluções.

Como referido anteriormente, a ESQM tem turmas maioritariamente de Ciências e Tecnologias e, nas disciplinas de Física e Química (11.º ano e Matemática (12.º ano), embora acima da média nacional, os resultados ficam aquém das nossas expectativas. É assim que no ano passado (em 12/13) encetamos um projeto, o **projeto Turma Alpha**, que tem como objetivos específicos diminuir as taxas de reprovação nessas duas disciplinas e melhorar as médias de escola em cada uma delas e, antes do mais, melhorar a aprendizagem dos alunos nessas disciplinas.

A *turma Alpha* funciona numa base de flexibilidade e rotatividade, em simultâneo com as turmas de origem, sem sobrecarga de horário para os alunos. Destina-se a fornecer as bases necessárias para fazer face ao exame aos alunos com défices no seu percurso, aulas para alunos com dificuldades de aprendizagem e um ensino mais personalizado e de apoio direto. Os conteúdos programáticos são os mesmos da turma original; não tem alunos fixos; é o professor da turma de origem que convida os alunos a frequentar a turma, sendo que os alunos não podem estar muito tempo afastados da turma de origem.

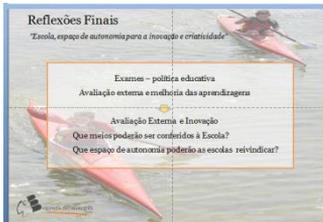
Temos tido bons resultados, mas falta ainda percorrer um longo caminho.

As constantes modificações na legislação, nos despachos de lançamento do ano letivo, os retrocessos dos processos de autonomia, a falta de

descentralização, a normalização excessiva, impedem muitas das soluções que gostaríamos de implementar.

Reflexões Finais

”Escola, espaço de autonomia para a inovação e criatividade”



Os exames fazem parte da política educativa, algo que foi colocado às escolas. Da Missão e Visão da escola ressalta que, considerando fundamentais os exames e os seus resultados, particularmente por nos fornecer um termo de comparação nacional para aquilo que pode ser avaliado nos exames (quer ao nível das classificações, quer ao nível das dificuldades dos alunos), a escola não está determinada por estes. Na ESQM não se ensina só para os exames, na medida em que se procura uma formação integral, mas não se descarta, de todo, o papel dos exames no ensino e na aprendizagem e não se descarta a sua importância no futuro dos alunos nas condições de acesso ao ensino superior e à profissão desejada. Mas se os exames não são o essencial, têm sido um ponto de partida para algumas das mudanças com que temos procurado melhorar o ensino e a aprendizagem na ESQM.

Os resultados dos exames também nos têm conduzido a uma reflexão constante. A ESQM tem tido um trabalho de perseverança, de reflexão e **autoavaliação**, de constante adaptação às mudanças e de implementar as ações necessárias para melhorar o seu desempenho também neste campo da avaliação externa. Assumimos as nossas responsabilidades. As escolas apropriam-se dos resultados, fazem o seu trabalho. Os exames deviam ser mais estáveis; é necessário um maior investimento nos critérios de correção e na calibragem dos exames. A tutela devia ouvir mais as escolas.

Tentámos referir os pontos principais do que a escola conseguiu fazer, do que lhe é devido e daquilo em que foi ajudada pela administração educativa ou por parceiros — o *Escxel*, por exemplo. Algumas coisas não foram conseguidas, por limitações próprias e de contexto — a

administração educativa e a legislação que produz, a falta de articulação com outras escolas, outros setores da administração.

A ESQM fez o seu percurso. Analisou os resultados. Mudou práticas. Conseguiu dar um salto qualitativo; atingiu novos patamares, está entre as melhores, mas sabe que tem ainda uma larga margem de progresso. Temos maior ambição. A escola gostaria de ir mais longe, ter uma **margem de autonomia** que lhe permita maior poder de decisão e de organização interna. É difícil transpor este patamar. Só é possível progredir com inovação, correndo riscos, com novos projetos e novas dinâmicas. É necessário uma verdadeira autonomia para encetar esse caminho; é necessário **confiança** nas escolas que já deram provas de desenvolver um bom trabalho.

A tutela tem de confiar nas escolas e nos seus diretores; necessitamos de uma administração menos “examinadora”, menos “normativa” e mais incentivadora de melhorias e mudanças, dinamizadora da diferenciação, do dinamismo e do potencial das escolas.

As escolas e a tutela deviam ter presente que todos pretendemos a melhoria do sistema educativo e dos seus impactos pessoais e sociais; termos a capacidade de abertura de nos escutarmos mutuamente, sabendo que rumamos todos na mesma direção; termos presente que **o sucesso de uns é o sucesso de outros e de todos**.

Esse pode ser um bom “exame” em que devemos passar, com garantidos bons resultados.

ENCERRAMENTO

José David Justino

O que vamos fazer com as conclusões deste seminário? É um bom desafio e eu não queria deixar de destacar duas ou três ideias.

Em primeiro lugar, agradeço a todos os que participaram e, acima de tudo, manifesto o meu regozijo em ver muitas escolas aqui. Não quer dizer que não haja mecanismos de representação dentro do Conselho Nacional de Educação, mas há uma coisa que devemos perceber muito claramente: é que este Conselho mais do que trabalhar para o Ministério tem de trabalhar para a Educação e para as escolas em Portugal. É nessa perspetiva que consideramos que proporcionar espaço para que estas escolas possam traduzir a sua opinião através da reflexão que fazem sobre as suas próprias experiências é muito positivo.

Para abordar o primeiro aspeto que julgo ser importante, depois de termos falado dos exames e das avaliações externas, gostaria de regressar às minhas tradições da História Económica e fazer aqui um contrafactual: E se não existissem exames, o que é que aconteceria? Esta é uma reflexão que vale a pena fazer porque, para todos os efeitos, não vi aqui nenhuma intervenção que dissesse: *O melhor é acabar como os exames, deixem-nos lá estar como estão, vamos voltar ao sistema antigo*. Ou seja, podemos concluir pela existência de um reconhecimento de que os exames têm um papel importante e de que a chamada avaliação externa tem esse mesmo papel importante. O problema que se põe é o de saber quais são as limitações e quais são os desafios que o atual modelo de exames coloca às escolas e às aprendizagens no sentido de podermos melhorar não só esse mesmo modelo mas o desempenho das aprendizagens na própria escola. E, nesse sentido, julgo que ouvimos aqui alguns contributos que de certeza não iremos esquecer. Ainda assim, colocou-se uma questão central para a qualidade das aprendizagens: Então e as avaliações internas? Ou seja, os muitos defeitos que podemos imputar às avaliações externas não se aplicarão às internas? Será que as avaliações internas não têm defeitos? O grande problema é que, durante muito tempo, as avaliações internas

funcionavam como uma espécie de *black box*, ou seja, não sabíamos o que é que se passava. Conhecíamos as classificações, mas como não havia nenhuma comparação possível entre o que é a avaliação interna e um referencial de avaliação externa — *é um 17 é um 17, é um 12 é um 12, é um 8 é um 8*. Com a introdução das chamadas provas de aferição, foi possível começar também a orientar um pouco as aprendizagens porque a dispersão de resultados na avaliação interna é muitíssimo maior do que a dispersão de resultados na avaliação externa. Isto é interessante. Por outro lado, se fizermos um gráfico de dispersão, detetamos que há escolas que propositadamente inflacionam as suas classificações. Especialmente no 12.º ano, que é isso que vai determinar se se entra ou não se entra no curso de medicina ou noutro curso qualquer. E, nesse sentido, não me interessam aqueles que estão na tendência; interessam-me os chamados *outliers*, aqueles que saem fora da tendência, que estão lá bem visíveis. Aqui a primeira sugestão que eu faria, eventualmente, ao Sr. Secretário de Estado é a de pôr a Inspeção de Educação a ver o que é que se passa em algumas destas escolas, porque o fundamental num sistema de avaliação externa, num sistema de exames, é a credibilidade. Se essa credibilidade não é assegurada através de um mecanismo de justiça, e há mecanismos de injustiça no acesso ao Ensino Superior, que é um bem superior, digamos assim, a credibilidade vai-se rapidamente. Claro que existe sempre uma margem de erro em avaliação, mas o problema é o erro sistemático e se analisarmos, ao longo dos vários anos, existem determinadas escolas que erram sempre no mesmo sentido e sempre na mesma altura. Portanto, seria interessante aprofundar esta questão, até para conferir essa credibilidade que um sistema de exames exige, pois, a partir da altura em que as pessoas não acreditarem que o sistema de exames é um instrumento equilibrado e justo, não obstante todas as limitações que possam existir, os exames deixam de ter um sentido, é a descredibilização total e deixa de haver qualquer preocupação com estas avaliações.

Mas tudo isso tem em grande parte a ver, também, com a avaliação interna. Valia a pena ponderar, e se calhar no futuro vamos fazê-lo, como é que se desenvolvem os diferentes modelos de avaliação interna e nas diferentes disciplinas. E isso tem em grande parte a ver com a capacidade,

ou, pelo menos, com as competências que os nossos professores tenham no que diz respeito à avaliação. É que se parte do princípio de que todos sabem avaliar e isso não é verdade. Eu, sinceramente, nos meus primeiros anos, quando era professor, cometi vários erros e tentei melhorar porque ia aprendendo com eles, mas ninguém me ensinou, nem ninguém me chamou à atenção para isso. Eu até teria agradecido que me tivessem chamado a atenção para estes erros. Ora dá a sensação que avaliar, nomeadamente em termos de avaliação interna, é um ato natural. Não é! É uma técnica própria, que exige determinado tipo de requisitos e determinado tipo de abordagem, sobre os quais julgo que valeria a pena pensarmos muito seriamente, porque creio que também é necessário fazer alguma formação contínua sobre modelos e técnicas de avaliação interna. Podemos apresentar aqui um manancial estatístico de ponderação, de aferição de itens, de tudo isso, mas depois as coisas esbarram naquilo que é a questão mais elementar, ou seja, porque é que eu dou 12 e não dou 14? Ou porque dou o 12 e não dou 8? Julgo que esse é um ponto sobre o qual vale a pena pensarmos.

Temos de olhar para os exames com algum cuidado e, acima de tudo, não nos podemos esquecer que os exames são no final de cada ciclo e que aquilo que nos interessa, em termos educacionais, é saber quais são os trajetos dos miúdos no que respeita à sua progressão e à forma como trabalham. Se nós apenas considerarmos a questão da avaliação no final de cada ciclo, fazemos análise de chamado *salto de barreiras*. Ou seja, era importante que pudéssemos ter uma corrida de 110 metros e não uma de 110 metros com não sei quantas barreiras. E o grande problema coloca-se muitas vezes no facto de essa avaliação não ter em conta os trajetos. E eu não sou grande defensor, embora reconheça que há muita gente que já trabalha sobre isso, do valor acrescentado, porque também entendo que o processo educativo não é análogo ao processo produtivo de uma fábrica. Enquanto com a economia nós conseguimos saber qual é a combinatória entre capital, terra e trabalho e, a partir daí, atribuímos preços e calculamos um valor final, na educação não é possível fazer isso, porque muitas das competências adquiridas só se refletem, em termos de valor acrescentado, muitos anos depois. Não sei, portanto, como é que se pode

fazer essa contabilidade e, sinceramente, penso que se deveria ter algum cuidado porque é muito interessante o conceito, mas a sua operacionalização levanta alguns problemas.

E agora recupero duas ou três frases que têm a ver com a comparabilidade dos exames. Na minha opinião, os exames não têm que ser comparáveis, mas era bom que fossem, ou seja, que fossem sujeitos a uma mesma métrica, um mesmo modelo. Por exemplo, tenho algumas dúvidas se não houve progressão nas aprendizagens e nos resultados como já hoje foi aqui afirmado. Embora seja sensível ao argumento que foi utilizado, continuo a achar que se andamos a fazer exames, e temos registos de exames, praticamente, há 14, 15 anos, e entretanto os testes internacionais revelam que nós progredimos nas aprendizagens — então como é que progredimos tendo por base os testes internacionais e não progredimos nos testes nacionais? São coisas diferentes, eu reconheço, mas não é possível compatibilizar isto? Ou seja, não pode haver aqui um paradoxo? Porque é que nós progredimos nos testes internacionais, com pequenas variações, e depois não o fazemos nos testes nacionais? Das duas uma, ou são os testes nacionais que não estão bem-feitos ou então, eventualmente, os testes nacionais obedecem a uma outra lógica que não responde a este problema da comparabilidade. De qualquer maneira, julgo que este tema é um tema sobre o qual valeria a pena refletirmos um pouco, e em conjunto, até mesmo para tentar apreender um pouco daquilo que é a experiência do IAVE porque, obviamente, quem está de fora pensa que sabe tudo, mas não sabe. Eu gosto sempre de ouvir quem está por dentro destas máquinas e destas instituições.

Nessa perspetiva, temos sempre o velho problema: para que é que servem os exames? Servem para avaliar aprendizagens ou programas? Porque as aprendizagens não são necessariamente aquilo que os programas apresentam e, portanto, temos aqui um triângulo entre avaliação externa, aprendizagens e programas, áreas que nem sempre coincidem. Agora, temos de dizer claramente o que é que queremos. Não podemos ter a pretensão de que o exame tenha que ir a tudo, não há hipótese. Qualquer avaliação externa é sempre uma avaliação limitada e, sendo limitada,

deve-se ter o maior cuidado com os resultados que decorrem dela. Mas também é necessário dizer que, sendo uma avaliação limitada, o problema é saber o que é que podemos tirar dela. Não é dizer que é limitada e mandá-la embora. Está bem, é limitada mas o que é que nos ensina? E este é um dos pontos importantes desta temática.

Relativamente aos *rankings*, estes são como as anedotas picantes. Uma boa anedota picante não diz asneiras. Porquê? Porque a maldade não está em quem faz o *ranking*, está em quem o vê. A anedota também é a mesma coisa, a maldade nunca está em quem conta a anedota, mas está em quem a ouve. E portanto, o problema não está na forma como se elaboram os *rankings*, porque é a mais simples, a mais simplificada, a mais linear que é possível, e, por isso, vale o que vale. Porém, há uma coisa que julgo ser importante: é que, e aliás o Presidente do Júri Nacional de Exames teve oportunidade de o salientar, só o facto de se ter publicado esses resultados alterou tudo. Para já, é um direito do cidadão: acesso à informação produzida pela administração pública. Em segundo lugar, quer queiramos quer não, essa publicação funcionou como um elemento de pressão sobre as culturas escolares e que foi aquilo que, de certa forma, a Diretora da Escola Secundária Quinta do Marquês disse hoje aqui. Ou seja, houve uma alteração das culturas escolares no sentido de dizer: *Não basta termos boas intenções, é necessário que as boas intenções produzam bons resultados*. Porque, durante muito tempo, andámos a ouvir que tínhamos os melhores professores, tínhamos os melhores alunos, tínhamos tudo isso... éramos os melhores. O problema foi quando publicaram os resultados internacionais. Foi um choque a publicação daqueles resultados do TIMSS... não sei se se lembram, em 95/96... foi um choque enorme. Afinal não somos tão inteligentes como nós pensávamos. Os alunos afinal não sabem tanto quanto nós julgávamos que eles sabiam. E é a partir daí que, na verdade, esta preocupação, de que não chega ter boas intenções, pois temos de ter bons resultados, vai obrigar as escolas a alterar a sua cultura, a forma como funcionam, o que não quer necessariamente dizer que tudo está focado no resultado... não, não pode ser... mas acima de tudo que temos que encontrar esse equilíbrio entre desenvolver boas aprendizagens e produzir bons resultados.

Penso que esta síntese reflete, de certa forma, entre muito mais que ficou registado e em que nós iremos trabalhar, algumas das minhas preocupações e creio que algumas delas podem ser facilmente partilháveis convosco. Nessa perspetiva, quero agradecer a vossa colaboração, prometo fazer do Conselho Nacional de Educação essa *casa comum da Educação* a que me referi no início deste seminário e fazer com que as escolas, os professores e os dirigentes possam encontrar aqui, neste Conselho, um ponto de reflexão conjunta e de partilha de experiências, que eu julgo que é o mais importante. E é a partir daí que nós podemos encontrar soluções que não sejam soluções de guerra, mas que acima de tudo sejam soluções de paz e de futuro.

E é nessa perspetiva que também ando a pensar que, eventualmente, precisemos sair desta casa. Ou seja, temos de fazer alguns seminários do CNE no Porto, na Covilhã, em Évora, em Faro, porque o CNE não pode estar confinado a estas paredes, embora muito agradáveis, mas tem que sair lá para fora, para poder também ouvir aqueles que provavelmente nem sempre têm hipótese de se poder deslocar e participar de forma ativa, como aqueles que aqui estão hoje fizeram.

Muito obrigado pela vossa presença, passo a palavra ao Sr. Secretário de Estado para encerrar os trabalhos.

João Grancho¹

Senhor Presidente do Conselho Nacional de Educação

Senhor Presidente do Instituto de Avaliação Educativa

Senhor Presidente do Júri Nacional de Exames

Senhoras e senhores Conselheiros

Começo por cumprimentar o Professor David Justino e aqui manifestar desde já o meu agrado pela sua designação para presidir ao Conselho Nacional da Educação e também pela forma como tem assumido as suas funções, procurando que esta casa de facto seja uma casa da educação, que ouça todos os intervenientes no processo educativo e que contribuam para a melhoria do sistema educativo.

Cumprimentar também as senhoras e os senhores conselheiros, os senhores representantes das várias instituições e associações, aqui presentes, o Sr. Presidente do IAVE, o Sr. Presidente do JNE, e dar nota de que a questão, o tema que foi escolhido para reflexão, para debate — “Avaliação Externa e Qualidade das Aprendizagens” — é um tema que percorre todo o sistema educativo e que preocupa naturalmente os Governos.

Todos os Governos têm uma perspetiva própria em relação à avaliação, às avaliações interna e externa e, nesse sentido, desenvolvem políticas que pensam serem as mais adequadas para que esse instrumento sirva para melhorar a qualidade das próprias políticas e, acima de tudo, sirva para melhorar os resultados escolares dos alunos.

Ora, nós sabemos do desenvolvimento a que estamos a assistir, em termos sociais e tecnológicos, e temos bem presente a pressão que é feita ou está assumida sobre todos os sistemas educativos.

¹ Secretário de Estado do Ensino Básico e Secundário.

Daqui decorre que aos sistemas educativos se exija que respondam de forma inovadora, adequada e eficaz às diferentes necessidades de cada momento e em função de cada contexto, com vista a assegurar uma educação de qualidade para todos. O mesmo é dizer: proporcionar as condições necessárias à aquisição e desenvolvimento de conhecimentos, capacidades e atitudes que permitam aos alunos tornarem-se cidadãos de sucesso.

Esse é o propósito de qualquer Governo e, naturalmente, é o propósito e objetivo deste Governo.

Sabemos também que, no quadro da OCDE, Portugal é um dos países que num curto espaço de tempo mais recuperou nos indicadores da educação.

No entanto, sabemos ainda que nos encontramos aquém das metas que pretendemos e devemos atingir. E à medida que vamos progredindo, maiores são os desafios e maior a dificuldade. E é assim, também, porque a escola se depara com uma pressão acrescida decorrente da heterogeneidade de públicos, de atores, de projetos e de expectativas.

Uma vez que está garantido e generalizado o acesso à educação, o desafio hoje está em acolher todos, educando e integrando socialmente cada aluno, desenvolver processos que permitam monitorizar e consolidar as aprendizagens dos alunos ao longo de todo o seu percurso escolar, com particular enfoque no 1.º ciclo enquanto alicerce fundamental de aprendizagem que permitirá enfrentar os ciclos subsequentes de uma forma mais sólida e com sucesso.

E neste contexto a avaliação surge como um processo indissociável do ensino e da aprendizagem, assumindo um papel relevante na melhoria da sua qualidade bem como da própria organização “escola”.

Não vou entrar na polémica dos *rankings*, porquanto é sabido que não é propósito deste Governo a sua especial valorização. Continuamos a disponibilizar os resultados acrescentando-lhes alguns elementos de contexto para que possam traduzir o mais fielmente aquilo que é a evolução desses próprios resultados. No entanto, relativamente à avaliação e à sua necessidade, a investigação vai-nos dando conta de que

governantes, gestores escolares, pais e alunos, todos estão interessados na avaliação e a vão utilizando, mais ou menos sistematicamente, de diversas formas, sendo certo que os Governos precisam de avaliar para estabelecer *padrões de excelência*, para monitorizar a qualidade da educação, para definir ou reformular políticas.

Como também referia o Professor Domingos Fernandes, “*As escolas e os respetivos órgãos de gestão utilizam a avaliação para identificar pontos fortes e pontos fracos dos seus projetos educativos, para planificar ou melhorar projetos e programas em curso ou para investir e intervir na gestão dos recursos humanos e materiais. Por sua vez, os professores e os diretores responsáveis pela vida das escolas utilizam a avaliação para monitorizar o progresso dos alunos, para avaliar o currículo e proceder ao seu ajustamento, para introduzir melhorias no processo de ensino, para identificar áreas problemáticas, para a diversificação de estratégias que contribuam para que todos os alunos possam aprender mais e melhor, e, sobretudo que saibam dar sentido à sua aprendizagem. Por fim, os pais e os alunos utilizam, e podem utilizar aliás, a avaliação para ajuizar acerca do trabalho realizado pelas escolas, conhecer e analisar as dificuldades evidenciadas, tomar decisões relativas ao prosseguimento de estudos, regular o processo de aprendizagem e também para a decisão sobre a escolha de escola*”.

A avaliação deve, pois, ser um instrumento que ajude a melhorar o ensino e a aprendizagem, a corrigir procedimentos, a integrar diferentes contextos, a regular e definir critérios transparentes que ajudem os alunos a organizar o seu estudo, motivando-os para a aprendizagem, a identificar dificuldades, permitindo que se possa intervir em diversos domínios.

Assim, as preocupações de monitorização constante devem resultar no desenvolvimento de esforços eficazes que contribuam para a melhoria dos resultados escolares.

Ao nosso nível, os relatórios produzidos pelas estruturas do MEC, designadamente a IGEC, o IAVE e o JNE, que tiveram aqui a oportunidade de já ter ouvido, no quadro da avaliação externa das escolas,

têm-nos dado o *feedback* relativamente ao modo como a avaliação externa dos alunos tem levado as escolas a repensar as suas práticas, a definir metas de aprendizagem com vista à melhoria dos seus resultados escolares, elaborando planos de melhoria envolvendo todos os intervenientes no processo educativo e formativo.

E, nesse sentido, é propósito do MEC contribuir para aumentar a eficácia do sistema de ensino português, através do desenvolvimento de mecanismos de garantia da qualidade e de novos padrões educacionais. Pretende-se dessa forma promover um sistema de educação equitativo e inclusivo que conduza a níveis de desempenho mais elevados por parte de todos os alunos, reduzindo as disparidades nos resultados escolares. Considera-se, nesse capítulo, fulcral o papel desempenhado pelas escolas na otimização dos processos educativos, adequando-os aos seus contextos específicos e às características particulares dos seus discentes, definindo, aplicando e avaliando estratégias educativas diversificadas de acordo com os recursos disponíveis com vista à promoção do sucesso escolar dos alunos.

Perante este quadro, que corresponde a um patamar de maior exigência relativamente à educação que temos e à educação que desejamos e que precisamos, levanta-se a questão de tentar perceber a evolução dos resultados de que temos vindo a tomar conhecimento. Nesse sentido, as várias estruturas do MEC têm vindo a enfatizar a necessidade de proceder à avaliação, à reflexão sobre esses mesmos resultados, no sentido de poder desenvolver estratégias de intervenção junto das escolas.

Nós sabemos que as escolas refletem sobre os resultados, mas o certo é que há regiões, há agrupamentos em concreto, em que sistematicamente não se vislumbram melhorias nos resultados escolares.

E aqui questiona-se o que pode e deve fazer a administração educativa. Numa perspetiva estrita de respeito pela autonomia das escolas, poderíamos dizer que cabe às escolas mobilizar os recursos e adotar as estratégias mais adequadas, cabendo ao Estado — ao Governo, ao Ministério da Educação —, confinar-se à verificação desses resultados.

Parece-nos que este não deve ser o caminho. Por isso temos estado a trabalhar, juntamente com o IAVE e com o Júri Nacional de Exames, exatamente no sentido de no próximo ano desencadearmos junto de um grupo restrito de escolas e de agrupamentos, que vão evidenciando resultados menos positivos, estratégias para ajudar a melhorar a sua capacidade de resposta. E isto significará envolver a administração educativa nos processos de melhoria, respeitando, naturalmente, a própria autonomia das escolas e não interferindo com aquilo que são diagnósticos entretanto já realizados.

Há, por certo, muitas outras questões que podem ser suscitadas em torno da avaliação. Aliás, o Sr. Presidente levantou aqui um conjunto de problemas, de questões, uma ou outra recomendação para o futuro próximo, mais ainda quando nos estamos a aproximar do período de provas e exames, designadamente uma divergência, um distanciamento entre resultados internos e resultados da avaliação externa. Porém, o que nós temos como propósito essencial é adotar, ao fim e ao cabo, aquilo que disse que seria necessário fazer: aproximar-nos do trabalho desenvolvido pelas escolas, contribuir para o reforço da sua capacitação para a melhoria dos resultados dos alunos e, neste plano, reforçar também a autonomia das escolas.

Aliás, muito recentemente alargamos a autonomia das escolas ao nível curricular. No plano curricular, 25% da carga curricular ou da carga horária pode ser decidida pelas escolas, que têm aqui uma margem muito grande de manobra. Mas isso tem necessariamente consequência, devendo-se ter em conta aquilo que se pretende que sejam os resultados que se venham a verificar depois com esse reforço de autonomia.

Vou concluir referindo que da parte do Ministério da Educação e Ciência há o propósito imediato de juntar, ao nível regional, tudo aquilo que são as boas experiências, as boas práticas das escolas que têm vindo a obter bons resultados, numa perspetiva colaborativa com as estruturas da administração educativa, no sentido da melhoria consistente dos resultados dos nossos alunos.

A todos agradeço esta disponibilidade para me escutar e ao Senhor Presidente do Conselho Nacional de Educação agradeço o convite que foi formulado.

Conselho Nacional de Educação

Rua Florbela Espanca

1700-195 Lisboa

Portugal

Tel.: (+351) 217 935 245

cnedu@cnedu.pt

www.cnedu.pt

