

Reflexão sobre Avaliação e a reforma do ECDU de 2009

Por decisão do Conselho Científico do IST foi colocada à discussão na escola um documento sobre os princípios regulamentadores de concursos, contratações e avaliações de desempenho. Este era um documento preliminar, que foi aprovado pelo CC do IST com o pressuposto de que era preciso abrir a discussão à escola, para que o tema fosse discutido de forma aberta e sem tabus. Não estavam, nem estão, tomadas decisões definitivas sobre este tema. Naturalmente, numa escola com a dimensão do IST só é possível ter uma discussão racional se existir um documento em cima da mesa para discutir.

Embora não estando envolvido na redacção do documento, como membro do CC tenho verificado uma grande abertura da direcção da escola a acolher modificações ao documento. Sei que a versão actual do documento, que ainda não foi publicitada, já difere bastante daquela que foi posta a discussão, registando a meu ver melhorias importantes e significativas. Não significa isto que eu esteja em total acordo com as soluções apontadas, pois subsistem opções de fundo com as quais eu não estou de acordo. Parece-me importante que continuem a surgir as contribuições e sugestões de alteração do documento.

Como membro eleito do CC, na discussão em sede de CC, tive oportunidade de manifestar as minhas objecções e de fazer sugestões de alterações ao documento. Na fase de discussão pública, tentei promover a discussão (e para isso organizei uma reunião da minha área científica para apreciação do documento) e procurei ouvir as opiniões de muitos colegas de forma a solidificar a minha posição. Estamos agora prestes a chegar a uma nova etapa da discussão. O documento voltará ao CC e espero que a nova versão do documento que sair dessa discussão volte ainda aos Departamentos.

Nesta fase parece-me importante partilhar as conclusões a que cheguei e explicitar a minha posição actual (até porque todos sabemos que “passar ao papel” impõe uma útil sistematização de ideias).

Na organização desta reflexão segui um método familiar para lidar com problemas, que inclui as seguintes etapas:

1. O contexto do problema
2. A identificação de soluções existentes para problemas análogos
3. A possibilidade de aplicação das soluções existentes ao nosso problema
4. Crítica das soluções já propostas
5. Sugestões de novas soluções

A literatura sobre este tema tende a ser extensa. As posições sobre estes assuntos costumam ser polémicas. Temo que esta reflexão sofra de ambos os males. Quem não quiser perder muito tempo com leituras e quiser descobrir quais são as alterações que eu proponho ao documento em discussão, poderá passar imediatamente aos pontos 4 e 5, que podem ser lidos de forma independente do resto do texto.

1. Contexto

Avaliar é uma das tarefas essenciais da carreira académica. A avaliação assume as mais diversas dimensões: a avaliação de alunos para aferir os seus conhecimentos, a avaliação de um artigo científico submetido para publicação, a avaliação dos resultados obtidos numa tese de doutoramento, a avaliação de investigadores que competem por um lugar, etc. A avaliação, qualquer que ela seja, pode não ser das tarefas mais gratificantes na carreira académica, mas sem ela(s) não é possível o progresso científico. Boas metodologias de avaliação podem contribuir significativamente para a qualidade de uma instituição, enquanto más metodologias de avaliação têm consequências nefastas a médio e longo prazo, muito difíceis de corrigir. As alterações de métodos de avaliação devem pois ser feitas com grande ponderação e bom senso, procurando corrigir os aspectos negativos que estão claramente identificados, potenciando o que há de bom no sistema, e evitando assumir riscos desnecessários.

A reforma do ECDU de 2009 colocou na ordem do dia o tema da avaliação dos docentes universitários nas suas várias dimensões: avaliação de desempenho, *tenure*, concursos e promoções. Trata-se de um tipo de avaliação em que o sistema universitário em Portugal tem registado falhas graves, de que são exemplos:

- Nunca ter existido qualquer tipo de avaliação de desempenho dos docentes universitários, sendo que até há poucos anos as subidas de escalão eram automáticas;
- O provimento definitivo nunca ter sido levado muito a sério, como se pode concluir do facto de ter sido concedido em 99,9% dos casos;
- Os concursos de recrutamento confundiram-se com as promoções, e regeram-se por regras contraditórias e limitativas que impedem, por exemplo, que os membros de um júri tenham um conhecimento efectivo das capacidades científicas e pedagógicas dos candidatos.

As falhas no sistema de avaliação foram ainda mais agravadas por uma total ausência de políticas de gestão dos quadros, de médio e longo prazo, nas instituições universitárias. No caso particular do IST existe ainda a questão da indefinição do modelo universitário Português, que oscila eternamente entre o equívoco de que “todas as universidades públicas são criadas iguais” e o das “instituições de referência”.

O prolongamento no tempo desta situação contribuiu para que se instalasse no sistema uma cultura de desconfiança, o que levou a que se começasse a recorrer com demasiada frequência aos tribunais. O sistema judicial (ele próprio com a sua conta de falhas graves) não contribuiu para a eliminação deste problema de confiança.

É neste contexto que se deve proceder à discussão do rumo a seguir e das alterações a efectuar na avaliação dos docentes no IST.

2. A avaliação noutros sistemas

Existindo por esse mundo fora sistemas universitários de qualidade manifestamente superior à do nosso sistema universitário, parece óbvio que se deva procurar nesses sistemas as soluções para os nossos problemas.

Apresento abaixo 4 exemplos típicos de como se processam os concursos, *tenures*, promoções e avaliações de desempenho. Este exemplos são apresentados de forma necessariamente superficial, mas aplicam-se em países europeus tais como a França, Holanda, Suíça, Dinamarca, etc.

Uma forma típica de organização de concursos de contratação é a seguinte:

- a) O Departamento abre um concurso numa ou várias áreas científicas e estabelece um comité de selecção;
- b) O comité de selecção recebe as candidaturas e forma uma “short list”.
- c) Os candidatos da “short list” são chamados para uma entrevista e/ou uma palestra. Simultaneamente, o comité de selecção procura obter opiniões de especialistas sobre os candidatos.
- d) O comité de selecção escolhe por consenso ou por votação a quem será feita a(s) oferta(s).

No caso de atribuição de *tenure*, os Departamento estabelecem regras que são do conhecimento público. Uma regra típica é a seguinte:

Criteria are conveniently organized under two headings: teaching and educational activities, and research. For promotion to tenure, each candidate is expected to be outstanding, or to show promise of becoming outstanding, in at least one of those two, and to be at least competent in the other category. A measure of outstanding performance requires demonstration of a high level of capability in contributing to, disseminating, or applying new knowledge, in a manner clearly recognizable by his or her peers. All candidates for promotion at all lower levels are expected to be at least competent in both categories, and show promise of meeting the criteria for promotion to Full Professor in due course.

e o procedimento típico é o seguinte:

- a) A *tenure* é atribuída após 3 ou 4 anos do início do contrato, e pode ser solicitada a qualquer momento pelo candidato.
- b) O comité de *tenure* do Departamento avalia o candidato e recomenda (ou não) a atribuição de *tenure*.
- c) O presidente do Departamento negocia com a administração da escola a atribuição da *tenure*, pois esta tem a última palavra a dizer.
- d) A atribuição de *tenure* é acompanhada pela promoção à categoria seguinte (e.g., de professor auxiliar a associado) e/ou mudança de escalão remuneratório.

As promoções de categoria são separadas dos concursos de contratação, e um procedimento típico é o seguinte:

- a) Um candidato que julgue encontrar-se em situação de ser promovido submete uma petição.
- b) Um comité de promoções do Departamento avalia o candidato e recomenda (ou não) a promoção do candidato.
- c) O presidente do Departamento negocia com a administração da escola a promoção, pois esta tem a última palavra a dizer.

Finalmente, a avaliação de desempenho é o aspecto da avaliação dos docentes que parece apresentar as maiores variações. Por exemplo, uma colega chamou-me a atenção para o sistema utilizado numa instituição dinamarquesa, que transcrevo:

The work of the researchers is not rated according to a fixed rating system. Instead, evaluation is done through discussions of the researchers' individual research plans and through evaluation of their research activities conducted through performance review meetings – the so-called “MUS” (administration and teaching) and “FUS” (research) meetings, which are held once a year between the researcher and the Head of department/Head of research.

Os critérios expectáveis para um professor sénior são especificados da seguinte forma:

We expect the following average research output per year per senior researcher employed full-time:

- 1 research publication per year
- 1 dissemination activity per year

In addition the senior researchers of the Department are also expected to continuously:

- * Participate in international, scientific conferences (with paper)
- * Participate in assessment committees
- * Act as referees and editors of scientific journals and books/book series
- * Help organise conferences, seminars, etc.
- * Participate in research networks
- * Participate in councils, boards, committees, etc.
- * Participate in applications on external funding

Many researchers are likely to have other activities and to publish other types of publications than the ones mentioned above, such as working papers, reviews or contributions to proceedings. Such publications are welcome, but are not given the same priority as publications which are regarded as ‘research publications’ and can in no way replace these.

3. Importar soluções propostas?

Porque não aplicar no IST modelos como os descritos acima? Será que modelos como estes não podem ser seguidos em Portugal? Infelizmente, a única resposta possível é “no contexto actual, não”.

No caso das contratações, em primeiro lugar o ECDU (versão 2009) proíbe explicitamente qualquer processo vagamente semelhante ao descrito acima, pois contém as seguintes disposições:

- Tem de existir um júri exclusivamente composto por membros da área científica em que é aberto o concurso.
- A maioria dos membros tem de ser exterior á universidade.

- Não está previsto a figura de “short list”. Todos os candidatos têm de ser tratados de forma igual e ordenados.
- Caso existam entrevistas, todos têm de ser entrevistados.

Em segundo lugar, atendendo à cultura instalada de recurso sistemático ao meio judicial, existem disposições de tribunais superiores que estabelecem que os critérios de ordenação e sistema de classificação têm de figurar nos editais dos concursos. Os júris têm também de ser publicados simultaneamente com os editais, não havendo lugar a consulta de especialistas sobre os candidatos presentes a concurso.

No caso das promoções, o ECDU (versão 2009) estabelece simplesmente que em Portugal estas não se podem distinguir das contratações, prolongando uma situação verdadeiramente inaceitável e que provoca uma enorme entropia no sistema.

O ECDU (versão 2009) reformulou todo o processo de atribuição de *provimento definitivo*, passando a fazer a distinção entre contratação por tempo indeterminado para professores auxiliares, e a *tenure* para professores associados e catedráticos. O ECDU estabelece que a *tenure* é atribuída no final de 1 ano de contrato, donde se conclui que, na prática, a decisão de atribuição de *tenure* é tomada no momento da decisão de contratação de um PAS ou PCA (módulo casos extremos). Por outro lado, no caso dos PAX a passagem a contrato por tempo indeterminado é o que se aproxima mais da *tenure* de outros sistemas, tal como descrita acima. Uma primeira diferença significativa é que o ECDU não permite, claro, que a escola decida a promoção automática a PAS. Tal como já acontecia anteriormente, o ECDU (versão 2009) dá uma grande liberdade à forma como a escola pode estabelecer este procedimento. Parece-me que este é um caso em que devemos corrigir os erros do passado e procurar adaptar as boas práticas internacionais relativas a *tenure*.

Finalmente, vamos passar a ter uma avaliação de desempenho que até agora não existia. O ECDU (versão 2009) dá uma grande liberdade à escola para estabelecer os seus mecanismos de avaliação e este é também um caso em que podemos importar soluções dos sistemas mais evoluídos. Mas, tal como em relação aos outros aspectos da avaliação, ao estabelecer o nosso modelo de avaliação de desempenho não devemos esquecer um factor importante: a cultura de desconfiança instalada no nosso sistema.

Apesar de todas as limitações à adaptação ao nosso sistema dos métodos de avaliação em voga nos sistemas mais evoluídos, há um princípio comum a todos eles que é possível (e importante) preservar:

A avaliação deve ser feita por pares e não existem fórmulas que os substituam.

No que segue, levarei em conta este princípio.

4. O PRCCAD

A actual gestão do IST avançou com uma proposta de método de avaliação, em todas as suas dimensões, baseada nos seguintes princípios fundamentais:

- Os concursos, contratações e avaliação de desempenho devem reger-se por um referencial único.
- A avaliação processa-se através da valoração de um conjunto de critérios que são o produto de factores qualitativos (factores discricionários) por factores quantitativos.
- O conjunto de critérios é agregado através de uma soma ponderada de onde resulta uma *classificação final quantitativa*.

O documento PRCCAD é de facto muito mais do que isto pois inclui uma listagem de todos os elementos que devem ser avaliados, procedimentos de tramitação e instrução para os vários aspectos dos concursos, contratações e avaliação de desempenho, etc. No seu conjunto representa um esforço notável de sistematização da avaliação, algo até aqui inexistente no IST.

Um dos aspectos que mais controvérsia tem levantado no PRCCAD é o do uso de fórmulas para se chegar a *classificações finais quantitativas* dos avaliados (sejam eles em concursos, contratações ou avaliação de desempenho).

Numa primeira fase do documento, que julgo ultrapassada, o princípio do referencial único era levado a um extremo tal que resultava numa *classificação final quantitativa* que servia para os diferentes tipos de avaliação. Felizmente, essa opção foi abandonada e passou-se a fazer uma diferenciação entre os diversos tipos de avaliação.

Um outro aspecto é que as fórmulas presentes no documento devem, de facto, ser consideradas como “não-fórmulas”, devido à presença dos factores discricionários. Essas “não-fórmulas” podem ser interpretadas como um método para condicionar ou delimitar o poder discricionário dos avaliadores ou dos membros dos júris, obrigando-os a justificar as suas decisões quando estas diferem de um determinado comportamento padrão.

Segundo os autores do documento, estaria desta forma resolvido dois dos principais problemas recorrentes nos recursos aos tribunais: a ausência de uma sistema de classificação e a falta de justificação de decisões. Embora não tenha certezas de que esses problemas serão resolvidos desta forma (e aqui confesso a minha total falta de preparação jurídica), vou aqui conceder que este tipo de condicionamento permite minimizar este problemas (e, portanto, permitir-me-ei utilizar a mesma filosofia numa das minhas propostas abaixo).

O facto de não existirem sistemas mais evoluídos (pelo menos que sejam do meu conhecimento) em que existam classificações finais quantitativas, semelhantes às propostas no documento PRCCAD, deve ser motivo, pelo menos, para uma reflexão (segundo, aliás, o princípio da prudência enunciado no primeiro parágrafo). Uma explicação plausível para esta inexistência é a de que não existe em mais nenhum país do mundo um ECDU como o nosso. No entanto, e apesar de concordar com o carácter verdadeiramente peculiar do nosso ECDU e os problemas criados pelos hábitos que se foram implantando, penso que há pelos menos dois tipos de objecções válidas a esta opção por uma *classificação final quantitativa*, e que me fazem sugerir alterações ao PRCCAD.

Em primeiro lugar, deve-se observar que muitos dos parâmetros que aferidos pelas fórmula são de tal forma subjectivos que os valores que se obtêm para os vários critérios têm muito pouca precisão. Se podemos contar artigos, determinar taxas de variação da intensidade de publicação, citações, factores de impacto, a qualidade de um artigo só pode ser aferida pelos pares, e é muito subjectiva. Na actividade pedagógica, medidas quantitativas são ainda mais difíceis de determinar. E o mesmo se pode dizer de todos os outros factores. A *classificação final quantitativa* que se obtêm ao agregar esses valores ainda menos precisão tem.

Em segundo lugar, há objecções que se prendem com os objectivos finais do método de avaliação em cada uma das suas dimensões e que estão em conflito com a existência de uma *classificação final quantitativa*. Nomeadamente:

- **Concursos:** O objectivo final de um concurso é o de seleccionar o melhor candidato. O facto de a lei (ou os tribunais) nos impor a classificação e ordenação de todos os candidatos, não deve ofuscar esse objectivo. A existência de uma *classificação final quantitativa*, por avaliador, não incentiva a que o júri funcione de forma colegial, debata as características dos candidatos e chegue a uma conclusão, de preferência unânime, sobre qual o melhor candidato. Aliás, uma decisão unânime do júri, com uma justificação comum, é provavelmente o mais eficaz desincentivador a um recurso aos tribunais. Em caso de grande desacordo entre os membros do júri é mais seguro não seleccionar ninguém.
- **Contratações por tempo indeterminado:** O objectivo final da avaliação é determinar se o candidato atingiu objectivos previamente especificados. Faz todo o sentido estabelecer critérios quantitativos mínimos (mas não suficientes) em certos vectores da avaliação, de forma a obter alguma objectividade na avaliação. Não vejo qual o objectivo de combinar critérios para obter uma *classificação final quantitativa*.
- **Avaliação de desempenho:** Uma vez estabelecido um número reduzido de patamares (e.g., insuficiente, suficiente, bom, muito bom) em que se processa a avaliação, interessa avaliar qual o patamar que o candidato atingiu. Uma vez identificados os elementos sujeitos a avaliação e estabelecidas as metas necessárias para se atingir cada patamar, não vejo a necessidade de uma *classificação final quantitativa* (mesmo no quadro de vários perfis de avaliação; esse podem ser diferenciados pelos patamares em cada parâmetro). O estabelecimento de uma *classificação final quantitativa* neste caso pode ser até elemento catalisador de um ambiente pouco saudável entre os docentes.

Em terceiro lugar, há um pressuposto em relação a todas as quantificações apresentadas no PRCCAD que não se verifica na prática: uma uniformidade dentro de uma mesma área científica. Pelo menos no caso da Matemática, nas áreas científicas existentes no DM essa uniformidade não se verifica.

Resulta de tudo isto, em particular, que a colocação de todos os coeficientes discricionários a 1,0 na fórmula para a *classificação final quantitativa* não representa nenhum comportamento padrão.

5. Propostas de alteração ao PRCCAD

Penso que o documento identifica correctamente todos os elementos sobre os quais se deve processar a avaliação. Interpreto o princípio de que “os concursos, contratações e avaliação de desempenho devem reger-se por um referencial único”, como afirmando que as várias formas de avaliação incidem de facto sobre a mesma actividade de docência e investigação, e por isso devem avaliar os mesmos elementos. Estou de acordo com este princípio genérico. As últimas versões do documento a que tive acesso apresentam várias simplificações e muitas correcções nas fórmulas, que elimina a maior parte dos problemas identificados na discussão pública do documento. A minha principal objecção é ao princípio de agregar critérios para obter uma *classificação final quantitativa*, mesmo contendo os tais factores discricionários, mesmo diferenciando as várias dimensões da avaliação. Em vez disso propõe-se:

Proposta 1. Os actuais factores quantitativos (otimizados) devem formar um *conjunto de indicadores* estruturados que devem ser fornecidos aos avaliadores (membros do júri, avaliadores de desempenho, comissões departamentais de contratação por tempo indeterminado) para (i) ajudá-los a formar a sua decisão (ii) estabelecer critérios mínimos, mas não suficientes, para admissão, contratação e avaliação.

Nota: O termo *indicador* não é aqui desprovido de significado. Na minha opinião, não existem “factores quantitativos perfeitos” que permitam avaliar de forma precisa um vector ou componente da avaliação. São sim meros indicadores que permitem aferir e sistematizar os vários aspectos da avaliação.

Em relação aos concursos, tal como referi acima, o objectivo deve ser determinar o melhor candidato. A ordenação dos candidatos é um factor secundário que não deve prejudicar o objectivo principal. Num júri com 5 elementos (o que deverá ser a regra no futuro, atendendo aos limites impostos pelo ECDU) uma decisão em que cada um dos membros do júri coloca em 1º lugar um candidato, sendo a ordenação decidida por fórmulas falíveis, não é aceitável. Assim, propõe-se:

Proposta 2. Os editais dos concursos devem conter o sistema de classificação (= pesos com que são valorizados os vários elementos de avaliação) bem como os critérios mínimos para admissão ao concurso, que podem ser quantificados pelos indicadores referidos acima (semelhante ao preconizado no PRCCAD). Os júris devem partir dos indicadores fornecidos e dos CV’s uniformizados, procurando chegar a decisões consensuais, incluindo uma justificação individual e uma ordenação dos candidatos comum a todos, devendo contar com apoio jurídico na elaboração da justificação. A falta de um consenso significativo em relação ao resultado do concurso deve ser motivo para o cancelamento do concurso. Tal como preconizado no PRCCAD, o presidente do júri desempenha aqui um papel essencial (devendo ser o presidente do Departamento).

Nos contratos por tempo indeterminado o principal objectivo deve ser o de contratar os docentes que demonstrem ter o potencial para ganharem um concurso de PAS a curto prazo. Para quebrar com a prática instalada é muito importante:

Proposta 3. Devem ser estabelecidos critérios quantitativos mínimos em relação a um subconjunto dos indicadores, necessários (mas não suficientes) para se obter a contratação por tempo indeterminado. Esses mínimos devem ser estabelecidos em cada Departamento, e por área científica. Admite-se que exista um período de transição e que estes apenas se apliquem às novas contratações de PAX. Estes critérios podem ser alterados no tempo, aplicando-se apenas às contratações feitas após cada alteração.

A avaliação de desempenho, sendo um processo que se repete no tempo de três em três anos, tem de ser necessariamente um processo simples e justo. Não se pode pretender com esta avaliação aferir a qualidade absoluta dos candidatos, mas sim a sua qualidade relativa e de forma aproximada, que permitirá premiar os melhores e penalizar os piores. O número de patamares da avaliação deve ser reduzido, permitindo distinguir 4 grandes grupos: insuficiente, suficiente, bom e muito bom. Seria expectável que a maioria dos docentes esteja no grupo “bom”, podendo dessa forma progredir na sua remuneração, a seu tempo. A classificação de “muito bom” seria reservada para premiar quem se distinguisse, sendo que a classificação de “suficiente” seria um sinal de que a performance do docente não está no “nível IST”. A classificação de “insuficiente” seria reservada para os casos problemáticos (a lei prevê que a atribuição da classificação mais baixa em dois períodos sucessivos origina um processo disciplinar).

Em relação à avaliação de desempenho, existe já um projecto de regulamento chapéu, ao nível da UTL, que obteve grande consenso entre os presidentes dos CC de todas as escolas, e que será discutido numa reunião do plenário do Senado da UTL marcada para o próximo dia 2 de Dezembro. Este regulamento aponta precisamente para este tipo de solução em 4 patamares. O que é pois necessário é que a escola crie uma grelha de avaliação que permite que cada avaliador coloque os seus avaliados em cada um desses patamares.

Proposta 4. Para efeitos de avaliação de desempenho devem ser criados 3 perfis a escolher pelo interessado: o perfil IST (incluindo os parâmetros investigação, docência, gestão e extensão universitária, com pesos apropriados), o perfil Docência (que exclui o parâmetro investigação) e o perfil Gestão (a que só se pode aceder excepcionalmente sob autorização). Os avaliadores, partindo de CV's uniformizados e dos indicadores, avaliam cada um desses parâmetros em quatro níveis (insuficiente, suficiente, bom e muito bom) e colocam os avaliados num dos 4 patamares, segundo uma grelha previamente estabelecida. Deve-se estabelecer as 9 horas semanais de docência como regra. Quem estiver no perfil “IST” e obtenha a classificação de “bom” ou “muito bom” vê a docência reduzida a 6 horas semanais.

Termino, repetindo o que disse acima: boas metodologias de avaliação podem contribuir significativamente para a qualidade de uma instituição, enquanto más metodologias de avaliação têm consequências nefastas a médio e longo prazo, muito difíceis de corrigir. Espero pois que a direcção da escola, tal como tem feito até aqui, continue a ouvir as muitas críticas e sugestões construtivas apresentadas ao documento PRCCAD, e que este documento evolua até obter uma adesão significativa por parte dos docentes do IST.

Rui Loja Fernandes
Departamento de Matemática,
IST 19 de Novembro de 2009