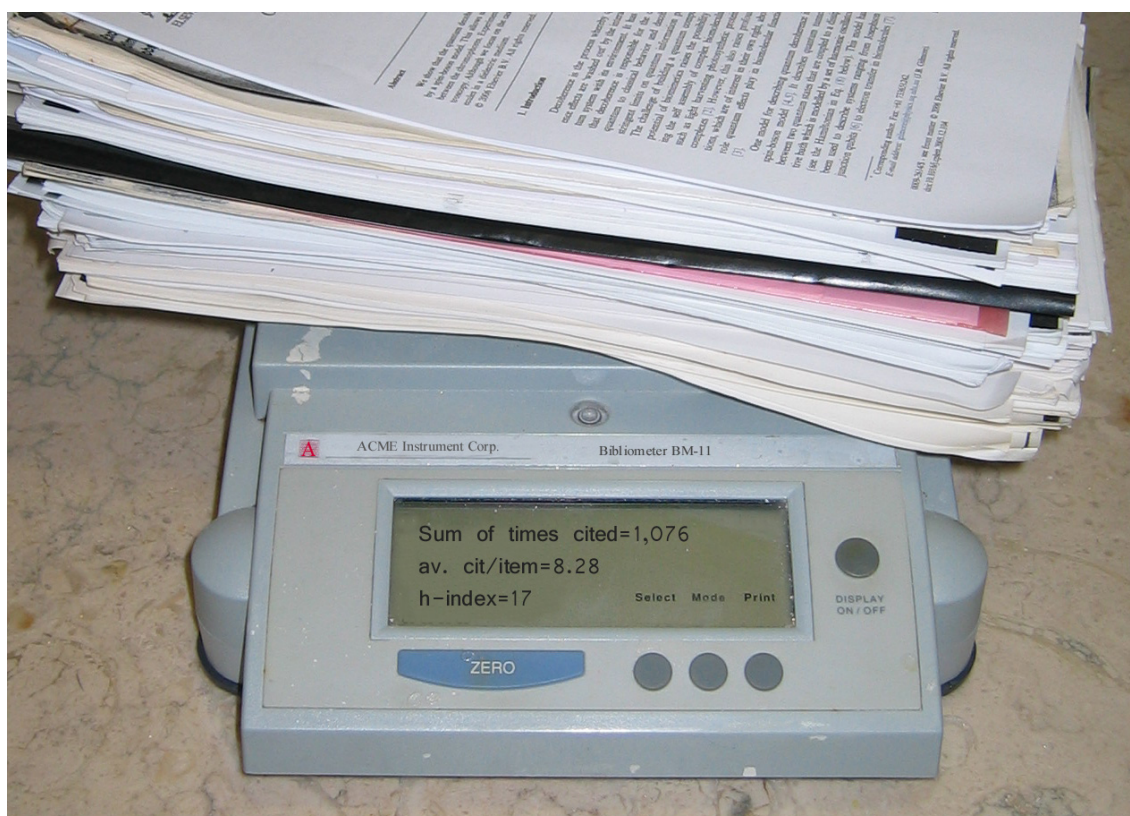


Algumas reflexões sobre avaliação em Ciência

Mário Berberan e Santos

*Centro de Química-Física Molecular, Instituto Superior Técnico, 1049-001 Lisboa,
berberan@ist.utl.pt*



1. Considerações gerais

A investigação científica tem desde há muito associadas várias formas de avaliação. Esta pode ser classificada quer de acordo com o seu *objecto*, quer de acordo com a sua *finalidade*.

Os objectos a avaliar são não só os *produtos de investigação* (artigos, livros, patentes, etc.) mas também os respectivos *produtores*. À primeira vista, estes são apenas os *investigadores individuais*. Na realidade, os investigadores não actuam completamente isolados, e são parte de estruturas complexas que condicionam a sua produtividade e impacto, pelo que são também produtores os *grupos*, as *unidades*, os *consórcios*, as *instituições*, os *países*, e até algumas *entidades supranacionais*. A avaliação pode ser feita a um só destes elementos (avaliação *individual*), ou a vários em simultâneo (avaliação de um *colectivo*), em que se determina a situação relativa de cada um.

A avaliação pode também ser classificada em *avaliação retrospectiva*, quando se examinam resultados obtidos, e em *avaliação prospectiva*, quando se analisa uma proposta de acções futuras. Em muitos casos, a avaliação é simultaneamente prospectiva e retrospectiva, uma vez que o desempenho passado diz muito sobre o possível desempenho futuro.

A avaliação pode ter por fim contribuir para o conhecimento de uma certa realidade, mas mais frequentemente destina-se a *fundamentar tomadas de decisão*, tais como a distribuição de verbas, a admissão ou possibilidade de progressão profissionais, a publicação de trabalhos em periódicos, e a atribuição de prémios.

Quando o objecto de avaliação é individual esta deve produzir uma resposta do tipo sim/não: publique-se ou não um artigo ou livro (avaliação de um produto), atribua-se ou não um doutoramento (avaliação de um investigador), etc. Eventualmente será também indicado o mérito em função de uma escala qualitativa, semi-quantitativa ou quantitativa (e.g. o antigo doutoramento com distinção e louvor, ou o artigo nos melhores 10%).

Quando o objecto é um colectivo, a avaliação é um processo que permite em primeiro lugar separar os elementos de um conjunto inicial (os candidatos) em dois conjuntos: os candidatos aprovados (em mérito absoluto) e os candidatos rejeitados. Mas isto não é normalmente suficiente: É ainda necessário ordenar os elementos aprovados segundo um dado critério que permita estabelecer o respectivo mérito relativo. É por exemplo com base nesta ordenação e nas suas disponibilidades

financeiras que a entidade financiadora decide quais os candidatos aprovados que irá apoiar, ou o apoio diferenciado que irá atribuir.

O processo de avaliação pode ser simples ou complexo, rápido ou demorado, transparente ou opaco, honesto ou desonesto. Estes atributos não são todos independentes. Um processo simples, rápido, transparente e desonesto seria por exemplo possível durante alguns períodos do Império Romano, mas não nos dias de hoje...

No cerne da avaliação está obviamente a entidade avaliadora. É concebível que em certas situações limite seja possível uma avaliação automática, baseada num algoritmo conhecido de antemão, dispensando-se a intervenção humana. Noutras, os avaliadores poderão ser totalmente livres nas suas escolhas. De um extremo ao outro, observam-se praticamente todos os casos.

A escolha do processo de avaliação é ditada pela sua finalidade. Deverá ser o mais simples, rápido e económico possível, por forma a atingir-se uma dada precisão nos resultados finais. Na avaliação de um colectivo com um grande número de elementos existirão sempre “falsos positivos” e “falsos negativos”. A sua eliminação total, ou quase, tornaria o processo demasiado lento e dispendioso. Este é um aspecto que os candidatos nem sempre compreendem, isto é, que é inevitável a existência de alguma contestação justificada dos resultados. A sua existência não implica que o processo de avaliação tenha sido incorrecto, pois uma resposta adequada (rápida e rigorosa) aos recursos faz parte do próprio processo, que não pode decorrer numa única etapa. A situação dos “falsos negativos” é assim corrigida com a possibilidade de recurso, desde que aqueles sejam em número muito inferior ao da população inicial, pois a não suceder tal o processo voltaria praticamente ao início. Já sobre os “falsos positivos” pouco há a fazer, uma vez que normalmente passam despercebidos. Será no entanto grave se resultarem de situações de favorecimento.

Em muitas situações há uma margem de incerteza inevitável. A utilização de um método automático de classificação só aparentemente resolveria o problema, pois a incerteza ou arbitrariedade já está presente nos dados de partida, dada a heterogeneidade dos candidatos.

Um processo de avaliação decorre incorrectamente se as suas regras não são cumpridas. Existe uma contradição interna óbvia. Em alguns concursos é possível deparar com justificações de voto injustificáveis, e até grosseiramente ilegais. Mas há formas mais subtis de distorcer intencionalmente os resultados. A escolha de um júri,

por exemplo, pode ser crucial. Também a adequação das regras de admissão ao fim pretendido tem sido um método de eleição, chegando-se por vezes a extremos caricatos, como sucedeu há alguns anos com um concurso denunciado pelos sindicatos do ensino superior, e que acabou por ser anulado. Mas os chamados “concursos com fotografia” são de tal forma comuns (isto é, ordinários) que têm uma designação bem estabelecida. A todos estes subterfúgios com que se procura esconder as verdadeiras intenções aplica-se a máxima do duque de La Rochefoucauld: *“l’hypocrisie est un hommage que le vice rend à la vertu.”*

2. Avaliação dos produtos de investigação

Um exemplo de avaliação individual de produtos de investigação é a avaliação de trabalhos submetidos a periódicos científicos para publicação, e cujo procedimento é bem conhecido (embora contestado por alguns). Pessoalmente considero-o de eficácia muito satisfatória, e dedico-lhe bastantes horas por ano como avaliador. As falhas que possa ter, e tem algumas, são em grande parte colmatadas pelo facto de os investigadores terem margem de escolha, podendo seleccionar a entidade avaliadora (o periódico em questão), e de poderem tentar várias vezes a publicação, se necessário. O acesso mais ou menos livre dos potenciais leitores às publicações faz o resto. De mencionar e saudar neste contexto as publicações de acesso aberto (*Open Access*), cujo número está em grande crescimento, mas cujo real impacto e consequências ainda não se podem aquilatar. Às iniciativas de acesso aberto têm respondido algumas grandes editoras quer com o acesso aberto a números antigos dos periódicos (por exemplo com mais de um ano), quer com o chamado *Free Access*, em que os autores podem pagar uma soma (elevada, da ordem de 2000 euros) para que o seu trabalho fique acessível a todos, normalmente como ficheiro pdf. Ao que se sabe, muito poucos autores têm optado por esta possibilidade. Mas o princípio, iniciado nos EUA, de que os trabalhos pagos com dinheiros públicos devem ter acesso público acabará por prevalecer, de uma forma ou de outra. A União Europeia também começa a evoluir neste sentido, tendo o seu Conselho Científico (ERC) já emitido algumas recomendações no final de 2006 [1].

A avaliação de um colectivo de produtos de investigação é menos vulgar, mas pode acontecer num concurso para um dado prémio (a melhor tese do ano numa dada área, etc).

3. Avaliação dos investigadores

À excepção de raríssimos investigadores com fortuna pessoal, todos os outros têm em princípio de se submeter a uma avaliação, periódica ou esporádica, como forma de justificar o apoio financeiro e material que recebem, ou querem receber, de entidades públicas e privadas. Infelizmente, o nosso sistema ainda permite que um “investigador” possa estar anos a fio sem produzir seja o que for, permanecendo numa condição profissional muito confortável. A avaliação científica determina (ou deveria determinar) também a progressão na carreira universitária, e, em geral, nas profissões com uma componente significativa de investigação.

3.1 Avaliações individuais de investigadores. Entre nós, a avaliação para progressão universitária é efectuada através de concursos (cada vez menos frequentes), e não por verificação de níveis de mérito atingidos individualmente. Como é bem sabido, não sucede assim noutros países mais desenvolvidos. No conhecido documento de Athans [2], aponta-se mesmo esta forma de progressão como um dos factores que impede as universidades portuguesas de atingirem níveis de excelência internacional. Escreve este autor: *“Assistant and associate professors must wait, for many, many years, for a “faculty opening” and a chance for promotion. More often than not, and in spite of a (ineffective) jury system, their promotion hinges on the whims of powerful “academic dictators” and depends on the past pattern of obedience and subservience to them. Academic dictators often manipulate the jury, especially since outside confidential recommendation letters are not commonly used. (.....) Excellence breeds excellence, while mediocrity fosters mediocrity.”*. Curiosamente, alguns dos mais antigos “tiranetes académicos” [3] (ditador seria demasiado forte e sério para o poder, exagerado e prejudicial, que efectivamente possuem ou possuíram no respectivo microcosmos) nunca passaram por uma verdadeira avaliação, tendo sido alçapremados a catedráticos por um decreto nocivo (pela forma como foi aplicado) que acompanhou a entrada em vigor, há quase três décadas, do Estatuto da Carreira Docente Universitária, estatuto esse que permanece aplicável salvo pequenas alterações. A inadequação da actual forma de progressão universitária é igualmente referida no relatório da OCDE sobre o Ensino Superior [4], cujas recomendações são claríssimas, indicando-se o que deve ser feito (obviamente por oposição ao que é feito): *“A big disincentive - especially for young Portuguese faculty members - is that promotions are based on openings rather than on merit. This together with the almost complete occupation of the full and associate professor positions can have serious consequences for the quality of research and*

teaching in many universities. The promotions and tenure decisions should be based solely on research and teaching achievement. Processes should be competitive, transparent and open; the evaluation of merit should be performed by outside peers in the field and preferably supported by recommendations from international experts.”

É lamentável que sucessivos governos que proclamam a sua dedicação à Ciência não tenham enfrentado o problema central dos mecanismos de progressão. Qual a explicação para esta paralisia crónica? A nova organização das universidades que se prepara irá alterar de facto a situação?

A avaliação do mérito individual é muito complexa, e tanto mais difícil quanto o campo do avaliado se afasta do campo do avaliador. Por essa razão os métodos bibliométricos têm grande utilização desde há décadas [5-7] (em Portugal desde há alguns anos, e apenas em algumas áreas). Para além de se considerar o número e a qualidade dos artigos publicados, medidos indirectamente (e nem sempre correctamente) através dos factores de impacto dos periódicos (problema dos falsos positivos), começou-se a dar há algum tempo grande atenção às citações recebidas. É de recordar neste âmbito o efémero (e único!) prémio da FCT “Estímulo à Excelência”, apresentado com grande alarde à comunicação social em 2004, e extinto apenas dois anos depois no mais profundo e comprometido silêncio. Felizmente, o acesso à *Web of Knowledge* não teve o mesmo fim.

A atenção dada às citações concentrou-se recentemente no índice h [8], um parâmetro numérico tentadoramente simples e já bastante usado (até porque foi adoptado pelo ISI em 2006). O índice h (o seu proponente, Hirsch, teve a modéstia de escolher uma letra minúscula) é sem dúvida útil, mas, tal como o número de citações, enferma de uma grande limitação: mede o impacto de *todos* os autores dos artigos em que um dado investigador participa. Por esta razão foi posteriormente proposto um índice h individual (divisão das citações de cada artigo pelo respectivo número de autores antes da aplicação do critério) que atenua este aspecto [9]. No entanto, fica ainda assim por determinar a posição na lista de autores (primeiro autor, autor responsável, autor secundário,...), e por ter em conta a escrita de livros, capítulos de livros, conferências, etc. De qualquer forma, verifica-se uma estreita correlação entre o índice h e o número total de citações, pelo que aquele parâmetro é em muitos casos redundante [9]. Note-se que os valores absolutos do número de citações e do índice h não têm o mesmo significado em todas as áreas científicas, pois são condicionados pelo

número médio de referências por artigo, que varia (segundo [10], este número é de 15 em Matemática e de 45 em Bioquímica).

Um investigador excepcional e com bastantes anos de actividade tem necessariamente muitas citações e um índice h elevado. Devemos contudo desconfiar de certos extremos numéricos: será por exemplo possível publicar 40 artigos num ano (situação verídica de um Prémio Nobel), com contribuição pessoal significativa em todos eles? Nestas 40 publicações incluem-se certamente situações de autoria honorária, em que são os co-autores a desejar a inclusão de um nome prestigiado (que não precisa de se preocupar com números de artigos ou com índices h). Não são aliás invulgares nos países desenvolvidos os grupos de investigação com várias dezenas de elementos, com equilíbrios internos mais ou menos delicados, organizados hierarquicamente sob um professor ou investigador notável, mas também bom gestor de recursos humanos e materiais, que vai apresentando em conferências os *highlights* dos trabalhos dos colaboradores, identificados em impressionantes fotografias de grupo que projecta no início ou no fim das mesmas. Afinal, até um “tiranete académico” pode reinar sobre um pequeno formigueiro de “obreiras científicas” em que algumas são mais capazes do que ele, e lhe permitem atingir um índice h razoável mas enganador [11].

Uma outra forma edificante de co-autoria honorária é a publicação cruzada, em que dois ou mais investigadores incluem os nomes uns dos outros nas respectivas publicações, num efeito multiplicativo surpreendente (não só sobre o número de publicações mas também sobre o número de citações), mas que nada tem de milagroso....

Não é necessário ser-se co-autor de muitos artigos por ano para se atingir (pelo menos nos últimos anos de carreira) um h elevado, digamos que superior a 30. Investigadores conceituados, trabalhando com grupos de pequena dimensão (contribuindo portanto de facto para os trabalhos), e publicando 3 a 5 artigos por ano, chegam a valores de h da ordem dos mencionados, ou mesmo superiores, em três a quatro décadas de actividade. Na área de Bioquímica, existe a elucidativa história sobre a produtividade de Gregorio Weber (1916-1997) [12]. Discutindo-se em reunião (ca. 1961) a sua possível contratação para o Departamento de Química da Universidade do Illinois (em Urbana-Champaign), e sendo objectado que, para cientista senior, tinha publicado um número pequeno de artigos [13], o presidente da Secção de Bioquímica confirmou-o, mas afirmou também que, coisa rara, o quociente entre o número de artigos notáveis e o número total de artigos era, no caso de Weber, igual à unidade.

Muito recentemente [14], foi proposto como melhor alternativa aos parâmetros acima indicados o número médio de citações por artigo (considerado fiável desde que o número total de artigos exceda 50), mas mais uma vez não se tem em conta o papel do investigador em cada publicação, e o parâmetro continua a ser função da área científica.

Deve ser sublinhado que a avaliação de um investigador baseada apenas num parâmetro numérico, seja ele h , o número de publicações, ou os números médio ou total de citações, é demasiado simplista. A avaliação pelos pares (recorrendo também aos índices bibliométricos [15]) continua a ser considerada o método mais fiável [5], *desde que* os pares sejam competentes e objectivos.

Há sempre que *avaliar os avaliadores*. Mas não basta assegurar a sua competência. É ainda necessário aplicar as condições de exclusão que o bom senso impõe, a União Europeia adoptou, e a prática nacional ignora, com as consequências perniciosas que se conhecem. Como exemplo das regras adoptadas na União Europeia, considere-se o seguinte extracto:

Conflicts of interest

*An expert involved in an evaluation must not have a **direct** or **indirect** conflict of interest with any of the proposals that they evaluate.*

*An evaluator is deemed to have a **direct** conflict of interest when any of the following applies: they are employed by the same institution and work in collaboration with the applicant at Department level; they work closely in collaboration with the applicant; they were involved in the preparation of the proposal; or they are in some other way closely related to the applicant (family relationship) or the work of the applicant (professional relationship) so as to compromise the evaluator's ability to impartially evaluate the proposal. **In such a case the evaluator should not take part in the evaluation of such a proposal and should not attend a panel meeting where such proposals are being evaluated.***

*An evaluator is deemed to have an **indirect** conflict of interest when none of the cases in the preceding paragraph applies and any of the following applies: the evaluator is employed by the same institution as the applicant; the evaluator would directly benefit from the proposal being funded or not funded in the context of their own research activities; the evaluator is involved in a contract or research collaboration with the applicant; or there is any other relationship with the proposal where the evaluator may not be able to impartially evaluate the proposal.(....)*

(European Commission, Human Resources and Mobility actions, *Guidance Notes for Evaluators*, 2003)

Este procedimento foi escrupulosamente seguido pela FCT nas últimas avaliações de projectos de Química e Bioquímica, mas é inexistente nos concursos

universitários, em que é vulgar existirem membros do júri com interesses directos em jogo. E assim se têm (de)formado Departamentos e Faculdades...

O mérito científico relativo é muito difícil de determinar, e por vezes talvez não possa mesmo ser estabelecido com rigor. Cada um terá ideias próprias a este respeito. Haverá quem dê menos valor ao trabalho de bancada, e haverá quem menospreze “os teóricos”. E dentro daquilo que nos atrai, tendemos a valorizar precisamente o que é mais difícil para nós.

O que irá ficar de duradouro de todas estas citações e de todos estes trabalhos? Muito pouco.... Basta pensarmos em tantos cientistas famosos, cujo nome está associado a um ou dois resultados. Mas ninguém trabalha para a posteridade, e a actividade de cada um tem muitas outras implicações para si e para os seus contemporâneos. Não se pode pois de forma alguma equiparar os que nada fazem aos que produzem. A relativização da actividade de cada um apenas o deve estimular, dando-lhe ao mesmo tempo a necessária humildade...

Será adequado insistir-se numa visão individualizante da Ciência, quando uma das suas características essenciais é a interacção? Se os resultados são obtidos por um grupo, é impossível dividi-los completamente pelos seus elementos. Uma situação extrema sucede na chamada *Big Science*, em que as publicações têm com frequência muitas dezenas de autores (reduzindo drasticamente os índices *h* individuais). Numa perspectiva de Estado/Sociedade, a melhor organização dos elementos será a que mais produzir, ao mais baixo custo. Haverá apenas que dar o devido reconhecimento aos elementos, consoante a contribuição para o esforço e resultado comuns, para que a máquina permaneça oleada. E é essa uma das finalidades da avaliação individual: estabelecer o que de facto é devido a cada investigador, a partir de resultados publicados em co-autoria [16].

3.2 Avaliações de colectivos de investigadores. É o caso dos concursos da carreira docente universitária. Nestes, é raríssimo que se apresente a concurso um único candidato, e quando isso sucede, é provavelmente mau sinal. Estes concursos implicam o estabelecimento de uma ordenação dos candidatos (mérito relativo), com rejeição prévia dos que não satisfizerem as exigências mínimas (reprovação em mérito absoluto). De acordo com a actual lei, para além do mérito científico é tido em conta o mérito pedagógico. Não é aqui o local para discutir em pormenor estes concursos, em que se viu já de tudo.

Nos concursos da FCT para bolsas de formação avançada (doutoramento e pós-doutoramento) avalia-se exclusivamente o mérito científico, quer dos candidatos, quer do programa de trabalhos, quer do responsável pela formação/instituição de acolhimento. O concurso é documental. Para o mérito dos candidatos utilizam-se algoritmos (de conhecimento público) que permitem uma classificação quase automática. No entanto a intervenção de um júri é também importante neste aspecto. A mera contagem de publicações prevista no algoritmo nem sempre é adequada, por exemplo. A avaliação não é obviamente perfeita, e as regras têm vindo a ser aperfeiçoadas (durante muito tempo, por exemplo, foram pedidas cartas de recomendação que não eram sequer lidas. A partir de agora deixam de ser necessárias). Há ainda que ter em conta a adaptação do sistema às regras de avaliação. Se é o número de publicações a ser considerado, há a tentação de associar o nome do candidato a bolseiro a trabalhos para os quais pouco ou nada contribuiu. Como saber qual a contribuição? Certas publicações exigem actualmente que a contribuição de cada autor seja explicitada. Mas isto é raro, e também sujeito a manipulação. O factor de impacto do periódico já é actualmente tido em conta, mas apenas qualitativamente, e deverá ser dividido pelo número de autores do artigo para ter maior significado. Será ainda de dar um maior peso ao trabalho do candidato quando este é primeiro autor. Tudo isto pode ser facilmente incorporado num algoritmo. Por outro lado, nem todas as áreas e sub-áreas têm factores de impacto comparáveis.

4. Conclusões

Foram abordados alguns tópicos relativos à avaliação em Ciência. Esta destina-se principalmente a fundamentar tomadas de decisão, sendo irrelevantes, para não dizer néscios, exercícios do tipo “os grandes investigadores”. Procurou-se estabelecer uma classificação dos diversos tipos de avaliação, de acordo com o objecto e a finalidade. Os objectos de avaliação são os produtos de investigação e os respectivos produtores. Os produtos de investigação típicos são os artigos, os livros, as teses e as patentes. Os produtores podem ser investigadores individuais, mas também grupos, unidades, consórcios, instituições, países e entidades supranacionais. A avaliação pode ser individual ou aplicada a colectivos.

Discutiu-se com maior pormenor a avaliação de artigos em periódicos e a avaliação de investigadores. Neste último caso foram referidas algumas patologias que

debilitam a investigação nacional, bem como a respectiva profilaxia. O mal está feito, e demorará décadas a desaparecer, mesmo que as medidas sejam introduzidas amanhã.

Outros assuntos importantes tais como a avaliação de projectos de investigação e a avaliação de entidades (grupos, instituições, etc.) serão abordados numa contribuição futura.

- [1] ERC statement on Open Access, <http://erc.europa.eu/pdf/open-access.pdf>
- [2] Athans M (2002) Portuguese research universities: Why not the best? *Economia Global e Gestão - Global Economics and Management Review* **7**, 121. O texto está disponível em vários *sites* nacionais.
- [3] É claro que muitos Professores Catedráticos não se enquadram nesta categoria definida por Athans.
- [4] Hasan A (2006) Tertiary Education in Portugal – Examiners’ Report, OCDE. O texto está disponível em vários *sites* nacionais.
- [5] Geisler E (2000) *The Metrics of Science and Technology*, Quorum, Westport.
- [6] <http://scientific.thomson.com/free/essays/>
- [7] Moed H F (2005) *Citation Analysis in Research Evaluation*, Springer, Dordrecht.
- [8] Hirsch J E (2005) An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci.* **102**, 16569. <http://www.pnas.org/cgi/content/short/102/46/16569>
- [9] <http://www.harzing.com/resources.htm#/pop.htm> e referências aqui citadas.
- [10] Garfield E (1997) Dispelling a few common myths about journal citation impacts. *The Scientist* **11** (3), 11. Disponível em <http://www.garfield.library.upenn.edu/>
- [11] Em tempos mais recuados falar-se-ia de bom grado num “condutor de homens”, hoje utiliza-se de preferência a forma mais refinada “capacidade de liderança”. Esta característica tem mesmo sido utilizada como critério de selecção (!) em alguns concursos universitários. A perspectiva implícita faz-me pensar na pergunta retórica de Estaline (um líder indiscutível) sobre o número de divisões do Papa.
- [12] Jameson D M (2001) The seminal contributions of Gregorio Weber to modern fluorescence spectroscopy. In *New Trends in Fluorescence Spectroscopy. Application to Chemical and Life Sciences* (Valeur B & Brochon J-C eds), Springer, Berlin.
- [13] Seriam aproximadamente 25 artigos. Durante toda a sua vida Weber publicou uns 170 artigos, vários capítulos de livro e um livro. Existem umas interessantes notas autobiográficas: Weber G (1989) Final words at Bocca di Magra. In *Fluorescent Biomolecules: Methodologies and Applications* (Jameson D M, Reinhart G D eds),

Springer, New York, livro que resultou de um simpósio em sua honra (1986) a que tive a sorte de assistir, ainda enquanto estudante de doutoramento.

[14] Lehman S, Jackson A D, Lautrup B E (2006) Measures for measures, *Nature* **444**, 1003. <http://www.nature.com/nature/journal/v444/n7122/full/4441003a.html>

[15] Garfield E (1987) Citation data is subtle stuff. A primer on evaluating a scientist's performance. *The Scientist* **1** (10), 9. Disponível em <http://www.garfield.library.upenn.edu/>

[16] Garfield E (1995) Giving credit only where it is due: The problem of defining authorship. *The Scientist* **9** (19), 13. Disponível em <http://www.garfield.library.upenn.edu/>

