

que esta dinâmica levou a uma disputa fratricida entre áreas do conhecimento, por exemplo, física e matemática *versus* ciências humanas e saúde coletiva, e atritos internos a cada um dos comitês, opondo pesquisadores que utilizam uma metodologia quantitativa àqueles que utilizam uma metodologia qualitativa. Essa tensão é desnecessária nos dias atuais, já que dispomos de várias abordagens metodológicas que permitem a síntese explícita de um componente subjetivo a um componente objetivo empírico, como o Bayesianismo, entre outras.

Por outro lado, se ainda assim quiséssemos argumentar favoravelmente a um retorno a um passado em que a avaliação da qualidade científica se dava em bases subjetivas, a quem caberia a escolha dos novos árbitros deste sistema? Se tomarmos como exemplo a dinâmica do nosso conselho de pesquisas, esta escolha se dá como fruto das ações lobistas das sociedades científicas e afins, entre outras influências. Como poderíamos garantir que esses novos árbitros assim escolhidos se tornariam os guardiões dos nossos anseios de alocação dos recursos da sociedade baseados numa atividade de pesquisa de qualidade? Não estaríamos apenas substituindo um problema por outro? Quem dentre nós não possui queixas ao sistema de *peer review*?

Se por um lado o conceito do que seja qualidade da produção científica me parece indefinido, a outra ponta da questão é igualmente ambígua, já que as supostas consequências imprevistas e indesejadas dos atuais processos avaliativos são de difícil caracterização. Há alguns anos, numa conversa informal com Eduardo Massad, professor titular da Faculdade de Medicina da Universidade de São Paulo, sobre este mesmo tema, uma observação nos chamou a atenção. Um exercício de avaliação do corpo docente daquela faculdade, utilizando diferentes métodos, levou a conclusões praticamente idênticas. Essa observação pode ser exemplificada no contexto do presente debate. Não seriam os cinco trabalhos mais importantes de sua produção, identificados pelo próprio autor como proposto por Bruce Alberts, também aqueles mais citados? Minha intuição (avaliação subjetiva) me instigaria a responder a essa questão com um sonoro sim, se eu não estivesse ciente das peças que nossa cognição nos prega.

Nosso sistema cognitivo possui muitas virtudes, mas não menosprezemos suas fragilidades. Kahneman <sup>1</sup> em seu livro reúne uma infinidade de exemplos que ilustram a nossa dificuldade em lidar com a ilusão da compreensão, e a ilusão de validade, assim como a nossa resistência em “domar” nossas intuições na presença de evidências empíricas. Avanços da psicologia cognitiva

aplicados a esse contexto parecem indicar que a questão não será resolvida contrapondo o subjetivo/qualitativo ao objetivo/quantitativo. A robustez e validade de nossas avaliações só parecem vislumbráveis se a realidade for apreendida por intermédio de uma síntese metodológica dessas abordagens. Não podemos abrir mão de nenhuma fonte de informação. Se utilizadas separadamente, essas avaliações serão ilusórias. A verdade emerge como uma interseção de mentiras <sup>2</sup>.

1. Kahneman D. Thinking, fast and slow. New York: Farrar, Straus and Giroux; 2013.
2. Levins R. Strategy of model building in population biology. Am Sci 1966; 54:421-31.

<http://dx.doi.org/10.1590/0102-311XCO050913>

*Luis Eugenio Portela  
Fernandes de Souza*

*Instituto de Saúde Coletiva,  
Universidade Federal da Bahia,  
Salvador, Brasil.  
luiseugenio@ufba.br*

## O desafio da avaliação da produção científica

A atividade científica institucionalizada adquiriu nas sociedades contemporâneas um estatuto de modo privilegiado de produção de conhecimentos válidos e úteis. Do ponto de vista do senso comum, são as aplicações técnicas resultantes da prática científica que parecem conferir esse privilégio e não considerações sobre seus pressupostos ontológicos, epistemológicos ou metodológicos <sup>1</sup>.

Contudo, o desenvolvimento da atividade científica, prévio às aplicações, requer a alocação de recursos pela sociedade. Além disso, a dinâmica da produção de conhecimento científico se caracteriza pelo questionamento permanente ao conhecimento estabelecido, seja ele comum ou científico <sup>2</sup>.

Nesse sentido, a avaliação da produção científica se impõe, tanto a partir de fora, ou seja, da sociedade, em geral, e dos governos ou das empresas que financiam as pesquisas, em particular, quanto a partir de dentro da própria comunidade científica que concorre pelos recursos, defendendo pontos de vistas diferentes ou mesmo contraditórios.

Vale fazer um parêntese para lembrar que submeter-se à avaliação não é uma especificidade da atividade científica. De uma forma ou de outra, toda atividade social é objeto de avaliação. Por exemplo, os políticos, em um regime democrático, são avaliados pelos eleitores, as empresas de capital aberto são avaliadas pelas

bolsas de valores, o desempenho econômico de um país é avaliado pelo PIB etc. Não é o caso de discuti-los aqui, mas registre-se que todos esses mecanismos de avaliação apresentam limitações significativas. Por que seria diferente com os cientistas?

Na área científica, a institucionalização da avaliação surgiu logo após a 2ª Guerra Mundial, podendo o *Relatório Bush* (1945) ser tomado como seu marco inicial. É importante destacar que, entre os cinco “valores fundamentais” que sustentavam a proposta de criação da National Research Foundation estava “...*assuring complete independence and freedom for the nature, scope, and methodology of research carried on in the institutions receiving public funds...*” (<http://www.nsf.gov/od/lpa/nsf50/vbush1945.htm>, acessado em 12/Jul/2013).

Esse valor – assim como os demais de modo menos direto – destaca a importância do respeito à autonomia dos cientistas na condução de seu trabalho, o que inclui o entendimento de que a avaliação da produção científica só pode e deve ser feita inter pares. Trata-se, de fato, de um privilégio que poucas ocupações laborais (as profissões, como vistas pela Sociologia) desfrutam. Assim, a comunidade científica tem tido a possibilidade de definir ela própria os mecanismos de avaliação a que aceita ser submetida, ainda que não possa simplesmente se recusar a se autoavaliar.

Os mecanismos de avaliação adotados evoluíram rapidamente. De início, utilizou-se o número de publicações como indicador da produtividade e do desempenho de cientistas e instituições de pesquisa. Em pouco tempo, entretanto, esse indicador foi abandonado e buscou-se alguma medida da qualidade das publicações. Considerando-se que só os próprios cientistas têm capacidade de avaliar um produto científico e que apenas os trabalhos de boa qualidade são efetivamente citados, adotou-se a prática de contar as citações. De novo, contudo, perceberam-se as falhas da simples contagem de citações ou mesmo da média de citações e, então, desenvolveram-se indicadores mais sofisticados, que relacionam o número de citações com o número de artigos citados, como o Fator de Impacto e o índice  $h^3$ .

Com a sua utilização, para diversos fins, os limites dos indicadores bibliométricos ficam mais e mais evidentes. A pretensão de avaliar a qualidade da produção científica por meio da relação entre número de documentos publicados e número de citações é prejudicada por fatores que os indicadores não podem captar, embora influenciem seus resultados: as diferenças da quantidade de documentos e da densidade de citações

por área do conhecimento, os ritmos distintos de obsolescência das publicações, as variações de escopo entre as bases bibliográficas e ainda os diferentes mecanismos de buscas, sem falar das distorções intencionalmente provocadas por várias formas de manipulação, como bem aponta Camargo Jr.

Nesse sentido, é um equívoco acreditar que a utilização dos indicadores bibliométricos assegura uma avaliação objetiva e absolutamente rigorosa do desempenho de pesquisadores e instituições científicas. Como também sinalizado no artigo em debate, acumulam-se evidências em contrário e, pior ainda, o aumento da frequência de utilização desses indicadores faz crescerem os vieses decorrentes de “pressões corruptoras”.

Permanece, portanto, o desafio de se encontrarem modos de avaliação que contemplem a complexidade do processo de produção de conhecimento, dando conta de identificar o que é, de fato, relevante.

A proposta apresentada por Camargo Jr. de se avaliar a produção científica com base no julgamento, por uma comissão de pares, de cinco produtos escolhidos pelos próprios pesquisadores ou instituições que se submetem à avaliação indica um caminho promissor de reforço da avaliação da qualidade da produção.

Como todo e qualquer procedimento avaliativo, no entanto, a proposta em si não oferece nenhuma garantia de avaliações justas e iguais, adotando as expressões da citada iniciativa *Research Excellence Framework*. Há vários aspectos que podem comprometer a justeza e a igualdade desse modo de avaliação: desde as idiosincrasias pessoais dos membros da comissão de avaliação até as disputas de poder em torno do capital científico, passando pelo acesso a fontes de financiamento.

Provavelmente, para evitar vieses, esse processo deverá ocorrer da forma mais transparente possível, ou seja, os avaliadores precisarão expor clara e publicamente os argumentos que sustentam sua decisão. A composição das comissões de avaliadores deverá ser cercada de muitos cuidados. (Quem avalia os avaliadores? Haverá a possibilidade de recurso a outra instância?). Enfim, há muitas questões em aberto e o autor da proposta faria bem em desenvolvê-la mais detalhadamente.

Isso dito, é necessário comentar alguns dos argumentos apresentados no artigo, incluindo o subentendido pelo título (ficção contábil?). Creio que os argumentos estão, fundamentalmente, corretos, mas talvez o estilo elegantemente provocador do autor o tenha levado a imprecisões.

A primeira das imprecisões se refere à oposição feita entre indicador quantitativo e avaliação

de qualidade. Ora, a utilização das citações, ainda que possa ser quantificada, é um indicador de qualidade, aceitando-se que a citação por outro cientista é um reconhecimento da relevância do produto científico publicado. Note-se que Camargo Jr. não a nega, ao contrário, considera a revisão por pares como o padrão-ouro da avaliação da ciência. Se é assim, a citação continua sendo uma medida válida, desde que se evitem exageros e distorções, como considerar que “duas citações indicariam um artigo duas vezes melhor do que artigos com apenas uma citação” (p. 1708).

A segunda imprecisão se atém à excessiva desvalorização da busca da objetividade. Concordo que não é questão de aceitar (pela fé?) a existência de uma “realidade povoada por objetos preexistentes que se oferecem passivamente aos nossos sentidos” (p. 1710), mas se trata de reconhecer o papel central da intersubjetividade nos processos de avaliação. Desse modo, fala a favor da sua adoção se os indicadores quantitativos facilitam a celebração de acordos intersubjetivos, o que não significa que quem os utiliza crê que a quantificação elimina toda a subjetividade.

A terceira está no título. Está claro que os indicadores bibliométricos estão sujeitos a vários tipos de distorção. Não são, contudo, uma ficção contábil. Com todos os seus limites, têm sentido: expressam, em certa medida, a produtividade e a qualidade do trabalho de pesquisadores e instituições.

Algumas medidas podem ser tentadas para reduzir as distorções: diversificar, por exemplo, as bases bibliográficas utilizadas para calculá-las (o Google Acadêmico vem sendo experimentado) ou utilizar os indicadores atribuídos aos artigos particularmente e não às revistas nas quais foram publicados.

Camargo Jr. está certo, de todo modo, ao sugerir não apenas que a avaliação da produção científica deve e pode ser melhorada, como também que isso passa pelo reforço da avaliação da qualidade. Sua proposta merece ser debatida e desenvolvida em todas as suas implicações. A avaliação, assumidamente subjetiva, de cinco trabalhos por uma comissão de pares é uma estratégia que, somada a outras, inclusive a utilização dos indicadores bibliométricos, pode tornar mais eficaz a identificação da produção científica relevante, e mais justos e iguais os processos avaliativos.

1. Santos BS. Introdução à ciência pós-moderna. 3a Ed. São Paulo: Edições Graal; 1989.
2. Bachelard G. A formação do espírito científico: contribuição para uma psicanálise do conhecimento. Rio de Janeiro: Contraponto; 1996.
3. Barreto M, Aragão E, Souza LEPF, Santana T, Barata RB. Diferenças entre as medidas do índice-h geradas em distintas fontes bibliográficas e engenhos de busca. Rev Saúde Pública; no prelo.

<http://dx.doi.org/10.1590/0102-311XCO060913>

Maurício L. Barreto

Instituto de Saúde Coletiva,  
Universidade Federal da Bahia,  
Salvador, Brasil.  
mauricio@ufba.br

### Como avaliar as ciências com uma deficiente ciência da avaliação científica?

Não existem muitas dúvidas sobre a importância da investigação científica como uma parte essencial da atividade humana e da atividade acadêmica, em especial. Os avanços do conhecimento e as inovações que gera são essenciais para nos fornecer o entendimento do mundo em que vivemos e para criar novas tecnologias que irão compor o modo de viver moderno. Entretanto, lado a lado com o grandioso e inequívoco espetáculo da ciência existe um cotidiano dos seus praticantes (pesquisadores) que inclui um ciclo complexo de atividades nem sempre fácil e, certamente, muito menos grandioso. No mundo acadêmico, esse ciclo passa pelo desenvolvimento de protocolos de investigação, busca dos meios necessários para implementá-lo, condução das reflexões/observações/experimentos necessários à sua concretização e por fim a redação e publicação, se possível em veículos qualificados, dos resultados encontrados. Um problema é que grande parte dessas atividades permanece em uma semiobscuridade, sendo na maioria das vezes apenas conhecida por um restrito número de investigadores do mesmo campo temático. Em outras palavras, grande parte dessas atividades que tanto mobiliza o mundo acadêmico é invisível para a sociedade e para os cidadãos que não são membros das comunidades científicas e, que, em última instância, as financiam. Para que sejam selecionados os indivíduos e instituições melhor qualificados e que serão selecionados e aquinhoados com os meios para executar os seus projetos, desenvolveu-se complexos sistemas de avaliação individual ou institucional de onde se pretende decantar a boa da má ciência e os bons dos não tão bons pesquisadores. Da mesma for-