

## Produção científica: avaliação da qualidade ou ficção contábil?

Scientific output: quality assessment or an accountant's tale?

Producción científica: ¿evaluación de la calidad o ficción contable?

*Kenneth Rochel de Camargo Jr. <sup>1</sup>*

<sup>1</sup> Centro Biomédico,  
Universidade do Estado do  
Rio de Janeiro, Rio de Janeiro,  
Brasil.

**Correspondência**  
K. R. Camargo Jr.  
Departamento de  
Planejamento e  
Administração em  
Saúde, Centro Biomédico,  
Universidade do Estado do  
Rio de Janeiro.  
Rua São Francisco Xavier  
524, 7º andar, Bloco D, Rio de  
Janeiro, RJ 20559-900, Brasil.  
kenneth@uerj.br

### Abstract

*Quality assessment of scientific output is based extensively (if not exclusively) on quantitative bibliometric indicators, despite mounting criticism. The reaction by the academic community has increased in recent years. After presenting a critical evaluation of this model, the article presents an alternative qualitative model oriented towards removing the excessive stimulus for the unbridled production of articles that make a dubious contribution to science. Criticism of the proposed model is presented and discussed.*

*Scientific Publication Indicators; Evaluation of Research Programs and Tools; Researcher Performance Evaluation Systems; Systems for Evaluation of Publications*

### Resumo

*A avaliação da qualidade da produção científica é baseada amplamente, quando não exclusivamente, em indicadores bibliométricos quantitativos, em que pesem as numerosas críticas aos mesmos. A reação da comunidade acadêmica vem se avolumando nos últimos anos. Após apresentar uma perspectiva crítica desse modelo de avaliação, expõe-se uma alternativa qualitativa, visando a remover o estímulo exagerado à produção desenfreada de artigos de contribuição duvidosa para a ciência. A crítica ao modelo proposto é apresentada e discutida.*

*Indicadores de Produção Científica; Avaliação de Programas e Instrumentos de Pesquisa; Sistemas de Créditos e Avaliação de Pesquisadores; Sistemas de Avaliação das Publicações*

## Introdução

Nos últimos dez anos, participei de diversas avaliações de pesquisadores e programas de pós-graduação. Cotejar essa experiência com o referencial teórico com que trabalho, os estudos sociais da ciência, me traz certo desconforto, que gostaria de compartilhar com a comunidade da qual faço parte e que procurei representar da melhor forma possível em tais avaliações. A avaliação da produção científica é indispensável, já que parte fundamental de seu financiamento provém de fundos públicos, especialmente no Brasil. Tanto decidir sobre a alocação de tais recursos como certificar-se de seu bom uso exigem alguma avaliação. Avaliar a ciência, porém, beira o impossível. Não há métricas adequadas para a mensuração do valor de uma pesquisa ou de seus produtos. Sua repercussão não é clara, nem previsível. A mecânica quântica, por exemplo, originalmente era pouco mais que um passatempo intelectual obscuro de físicos, mas podemos dizer sem exagero que o mundo que habitamos foi criado com linhagens tecnológicas diretamente retraçáveis às pesquisas de Max Plank, pioneiro do campo. Sem modalidades óbvias de mensuração, criamos indicadores diversos, discutíveis mas beneficiando-se da ilusão de “objetividade” conferida pela expressão em números, mesmo não tendo significado claro. Seria redundante reiterar críticas já feitas às fragilidades dos indicadores bibliométricos usuais<sup>1</sup>; muito resumidamente, são médias, indicadores problemáticos para distribuições tão irregulares como a de citações, e baseiam-se na premissa fundamental e não discutida de que citações não apenas indicam qualidade como o fazem de modo cardinal, ou seja, duas citações indicariam um artigo duas vezes melhor do que artigos com apenas uma citação, além da existência de inúmeros problemas metodológicos no seu cálculo. A ênfase na suposta “objetividade” de indicadores numéricos leva a que não se questione o aspecto fundamental da utilização de tais indicadores, seu sentido<sup>2</sup>. Por outro lado, estudos que procuram entender o processo de inclusão de referências bibliográficas em artigos não dão respaldo a nenhuma dessas suposições<sup>3</sup>. Acrescente-se a isso o fato de que muitas revistas científicas limitam o número de referências por artigo, o que encoraja autores, por exemplo, a citar revisões ao invés de buscar referências originais, criando problemas adicionais de atribuição de crédito<sup>4</sup>. As reações aos problemas causados pela utilização de tais indicadores vêm se avolumando. O processo avaliativo em curso no Reino Unido, *Research Excellence Framework* (REF), na área de respostas para perguntas frequentes (FAQ) em sua página de

Internet (<http://www.ref.ac.uk/faq/researchoutputsref2/>, acessado em 27/Abr/2012) responde da seguinte forma à pergunta sobre como indicadores de impacto serão utilizados na avaliação: “*Nenhum subcomitê fará uso de ranqueamentos, listas ou fatores de impacto de revistas ou da percepção da relevância de editoras na avaliação da produção da pesquisa. Um princípio subjacente ao REF é que todos os tipos de pesquisa e todos os tipos de produção de pesquisas abarcadas por todas as disciplinas devem ser avaliadas em bases justas e iguais*”. Veja-se ainda o manifesto *San Francisco Declaration on Research Assessment* (DORA; <http://am.ascb.org/dora/>, acessado em 20/Mai/2013), de associações e pesquisadores propondo a eliminação do uso de métricas de citação em avaliações da produção científica.

## Uma perspectiva crítica

A crítica que sustento, porém, vai além desses instrumentos; a questão fundamental é que substituir a avaliação da *qualidade* da ciência por qualquer avaliação baseada predominantemente em indicadores quantitativos é em si problemática, no mínimo. Talvez uma analogia, em que pesem todas as limitações de uma argumentação baseada nela, expresse melhor os absurdos ocultos pela numerologia supostamente científica. A Arte é tão ou mais difícil de avaliar que a Ciência. Suponhamos que alguém desejasse ranquear “objetivamente” pintores segundo a sua “produtividade”. Um indicador óbvio seria o número de quadros, mas nem todos os quadros são iguais. Seria possível medir cada um e determinar a sua área, chegando a um indicador padronizado, que, para permitir comparações, poderia ser dividido pelo período de atividade do pintor. Poderíamos medir o tempo que apreciadores levam contemplando cada quadro, e seria possível estabelecer ainda uma relação de museus segundo a sua importância, levando em conta o número de visitantes. Melhor ainda, poderia ser computado o número de vezes que estudantes de artes fazem cópias dos quadros, o que daria uma medida da importância atribuída pelos próprios artistas à obra de cada pintor. Finalmente, o número de prêmios e críticas favoráveis recebidos por quadros e pintores poderia ser outro indicador. Ao final desse processo teríamos um conjunto de indicadores precisos, objetivos e facilmente reprodutíveis, segundo os quais Ticiano provavelmente seria um pintor mais produtivo que da Vinci. É provável que Bosch e da Vinci estivessem fora de listas de pintores mais produtivos, e talvez pintores relativamente medianos, mas com produção regular, se destacassem no conjunto.

Artistas que tinham estúdios, como Rembrandt, se beneficiariam da autoria de trabalhos coletivos. A produção inicial dos impressionistas, porém, seria desqualificada, já que foi extremamente mal recebida pelos críticos – o próprio rótulo “impressionista” foi usado inicialmente de modo pejorativo. Poderíamos ainda concluir que apenas o Louvre e a Tate Gallery de fato contam como museus relevantes, orientando amantes da arte a não perder tempo com museus menores como L’Orangerie. A questão que caberia perguntar ao fim desse exaustivo exercício é se de fato haveria algum sentido nessas conclusões. É óbvio que se trata de uma analogia, por isso mesmo imperfeita, mas estaríamos de fato fazendo algo radicalmente diferente da caricatura que propus?

### Consequências imprevistas (e indesejadas)

Nos congratulamos, com razão, pelos resultados positivos dos processos avaliativos a que nos submetemos, mas não temos considerado suas consequências imprevistas – e eventualmente indesejadas. Um cientista social americano, Donald Campbell, ao estudar as avaliações de intervenções sociais <sup>5</sup> (p. 49), formulou a seguinte observação, conhecida como *Lei de Campbell*: “Quanto mais um indicador social quantitativo for usado em processos sociais de tomada de decisão, mais sujeito será a pressões corruptoras e mais apto estará a distorcer e corromper os processos sociais que se pretende monitorar com seu uso”. Alguns estudos já apontam para evidências da ação deletéria de avaliações que pressionam pela quantidade; Fanelli, por exemplo, demonstra que tal pressão aumenta o viés de publicação de resultados positivos <sup>6</sup>, enquanto que Fang et al. <sup>7</sup> mostraram que a maior parte das retratações de artigos publicados em revistas biomédicas e das ciências da saúde é devida à má conduta na pesquisa, e que a retratação por fraude ou erro é mais frequente justamente nos periódicos de maior impacto. Face à pressão competitiva de publicar cada vez mais, multiplicam-se subterfúgios para incrementar a quantidade de artigos e capítulos publicados, e mesmo de citações. A produção em série de artigos sem maior interesse ou inovação, ainda que fundamentalmente “corretos”, a multiplicação do número de autores para cada texto sem que se considere adequadamente a atribuição de autoria, a distribuição de conteúdos entre diversas publicações (conhecida como “publicação salame”) e a criação de verdadeiros clubes de citação mútua são fenômenos há muito conhecidos na literatura mundial e que começam a se manifestar em nosso meio. Coincidentemente, a publicação salame

foi o tema da coluna recente de Fernando Reinach, professor titular da Universidade de São Paulo <sup>8</sup>. Reinach chega a conclusões semelhantes a Campbell. Destaco de sua análise um trecho que me incomodou, em particular: “*Não há dúvida de que métodos quantitativos são úteis para avaliar um cientista, mas usá-los de modo exclusivo, abdicando da capacidade subjetiva de identificar pessoas talentosas, criativas ou simplesmente geniais, é caminho seguro para excluir da carreira científica as poucas pessoas que realmente podem fazer descobertas importantes. Essa atitude isenta os responsáveis de tomar e defender decisões. É a covardia intelectual escondida por trás de algoritmos matemáticos*”. Estaríamos (eu inclusive) de fato sendo covardes intelectuais? No mínimo não temos mostrado ousadia na proposição de inovações em nossos processos avaliativos que considerem efetivamente a qualidade da produção. Soçobrando num mar de textos que nós mesmos incentivamos a crescer exponencialmente, que provavelmente contribui mais para a deterioração da relação sinal/ruído da comunicação científica do que para a inovação, invocamos o gigantismo por nós mesmos incentivado como limitação para que de fato se avalie qualidade – tentar ler o que está sendo publicado corresponde à proverbial expressão do inglês *drink from the firehose* (beber da mangueira de incêndio).

### Uma alternativa

As estratégias de avaliação em uso estimulam a publicação de volumes crescentes, com os efeitos deletérios já apontados. Para quebrar esse ciclo, será necessário uma ruptura radical com a tradição de ranqueamentos segundo indicadores quantitativos de produção, e agregar, tal como já se faz em vários países e como foi recomendado pelo Comitê Assessor de Saúde Coletiva e Nutrição do CNPq (do qual fiz parte), um pequeno conjunto de produtos (não necessariamente publicações) ligados ao projeto ou programa que possam ser examinados, para que a elusiva qualidade da produção seja de fato considerada nas avaliações. Isso é defendido, por exemplo, por Bruce Alberts, pesquisador, ex-presidente da National Academy of Sciences (Estados Unidos) e editor da *Science* <sup>9</sup>: “*Eu apoio o sistema usado nos Estados Unidos. Quando um pesquisador está pronto para ser promovido ou próximo de ser avaliado, pede-se a ele que separe os cinco trabalhos mais importantes que produziu. Essa é a sua contribuição para a ciência. Ninguém pede todo o trabalho do candidato. Desse modo eu posso ler os cinco itens do seu trabalho e avaliar o candidato. Mas é impossível fazer isso se eles me dão 80 arti-*

gos. Nesse caso, tudo o que posso fazer é ver onde os artigos foram publicados". Além dos Estados Unidos, citados por Alberts, outros países adotam essa lógica de avaliação, como o Canadá e o Reino Unido<sup>4</sup>. A reação negativa a esse modelo repousa basicamente no argumento de que dessa forma a avaliação deixaria de ser "objetiva", tendo implícita a ideia de que estaria sujeita a manipulações de diversas ordens. É aqui que a crítica epistemológica se encontra com a de Reinach; por um lado, a ilusão fundada numa variante ingênua de realismo de que "qualidade científica" é algo com existência própria, mensurável por indicadores bibliográficos. Por outro, o medo da tal "subjetividade" justificando a "covardia intelectual" a escondida por trás dos algoritmos. Uma exploração a contento desse tema não é possível neste espaço, mas é necessário ressaltar ao menos alguns aspectos. Em primeiro lugar, "objetividade" não é um conceito autoevidente; com frequência o que está por trás dessa palavra é apenas uma noção sem maior elaboração reflexiva, fundada numa visão de mundo ultraempírica, arcaica, de uma realidade povoada por objetos preexistentes que se oferecem passivamente aos nossos sentidos, sendo a mente humana apenas um "espelho da Natureza", na expressão de Rorty<sup>10,11</sup>. Para Popper, por outro lado, "objetividade" seria a possibilidade de produção de acordos intersubjetivos no marco da metodologia da ciência<sup>12</sup>, o que levou Hacking a formular o conceito de tecnologia de intersubjetividade<sup>13</sup>. Medidas quantitativas podem ser consideradas "mais objetivas" nessa lógica por conduzirem mais facilmente a acordos intersubjetivos, mas não por serem ontologicamente "mais reais", se é que esta última expressão tem algum sentido, filosoficamente falando. No caso em questão, tendo em vista que o nível mais elementar de onde se pode obter alguma informação – o artigo publicado – depende fundamentalmente de um processo qualitativo e altamente subjetivo, que é a revisão por pares, chega a ser surpreendente a crença de que a partir daí é possível eliminar-se toda a subjetividade, desde

que se quantifique alguma coisa. O que se propõe com a reintrodução de uma lógica qualitativa na avaliação, portanto, é simplesmente que se recoloque o secular padrão-ouro da avaliação da ciência, a revisão por pares, em seu devido lugar. A solução para o problema da possível manipulação não é refugiar-se em algoritmos que decidam automaticamente desonerando-nos desta responsabilidade, até porque tampouco eles são infensos à manipulação, mas a plena transparência dos processos de avaliação.

## Conclusão

A efetiva implementação de uma sistemática qualitativa de avaliação no Brasil certamente demandará mais tempo do que o exíguo prazo de que se dispõe atualmente, e deverá contar com a colaboração da comunidade acadêmica, elaborando pareceres adequados que de fato subsidiem o trabalho dos comitês de avaliação. A Presidência do CNPq sinalizou recentemente para a adoção desse processo; cabe à comunidade científica cobrar sua implantação e disseminação para outras instituições. Que não se pense, contudo, que há aqui uma recusa ao estabelecimento, por exemplo, de metas mínimas, pactuadas de produção; não se trata de criar justificativas para o laxismo de pesquisadores que devem sim ser incentivados, e mesmo cobrados, a publicar. O que se critica aqui é a noção prevalente de que é possível avaliar a qualidade da produção científica apenas ou principalmente por meios quantitativos, e o clima de "publique-se a qualquer preço" que essa lógica encoraja. Este artigo inicia com uma referência ao desconforto com o *status quo*; se tal desconforto, e mesmo descontentamento é compartilhado por outros pesquisadores, cabe perguntar-nos até quando nos sujeitaremos passivamente a processos de normatização de nossa atividade que percebemos não apenas como inadequados, mas mesmo danosos?

## Resumen

*La evaluación de la calidad de la producción científica se basa ampliamente, si no exclusivamente, en los indicadores bibliométricos cuantitativos, a pesar de las numerosas críticas a los mismos. La reacción de la comunidad académica se ha acumulado en los últimos años. Después de presentar una evaluación crítica de este modelo, una alternativa cualitativa es expuesta, con el objetivo de eliminar el estímulo excesivo a la*

*producción desenfrenada de artículos de dudosa contribución a la ciencia. Se presenta y se discute la crítica al modelo propuesto.*

*Indicadores de Producción Científica; Evaluación de Programas e Instrumentos de Investigación; Sistemas de Créditos y Evaluación de Investigadores; Sistemas de Evaluación de las Publicaciones*

## Referências

1. Camargo Jr. KR. O rei está nu, mas segue impávido: os abusos da bibliometria na avaliação da ciência. *Saúde & Transformação Social* 2010; 1:3-8.
2. Adler R, Ewing J, Taylor P. Citation statistics – a report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS). *Stat Sci* 2009; 24:1-14.
3. Krampen G, Becker R, Wahner U, Montada L. On the validity of citation counting in science evaluation: content analyses of references and citations in psychological publications. *Scientometrics* 2007; 71:191-202.
4. Lawrence P. Lost in publication: how measurement harms science. *Ethics Sci Environ Polit* 2008; 8:9-11.
5. Campbell DT. *Assessing the impact of planned social change*. Hanover: The Public Affairs Center, Dartmouth College; 1976.
6. Fanelli D. Do pressures to publish increase scientists' bias? An empirical support from US states data. *PLoS One* 2010; 5:e10271.
7. Fang FC, Steen RG, Casadevall A. Misconduct accounts for the majority of retracted scientific publications. *Proc Natl Acad Sci U S A* 2012; 109: 17028-33.
8. Reinach F. Darwin e a prática da "Salami Science". <http://www.estadao.com.br/noticias/impresso,darwin-e-a-pratica-da-salami-science-,1026037,0.htm> (acessado em 27/Abr/2012).
9. Alberts B. We need to teach science. <http://revista-pesquisa.fapesp.br/en/2012/09/10/bruce-alberts-we-need-to-teach-science/> (acessado em 29/Mai/2013).
10. Rorty R. *Objectivity, relativism, and truth*. Cambridge: Cambridge University Press; 1991.
11. Rorty R. *Philosophy and the mirror of nature*. Princeton: Princeton University Press; 2009.
12. Popper K. The sociology of knowledge. In: Stehr N, Grundmann R, editors. *Knowledge: critical concepts*. V: sociology of knowledge and science. New York: Routledge; 2005. p. 63-74.
13. Hacking I. Statistical language, statistical truth, and statistical reason: the self authentication of a style of scientific reasoning. In: McMullin E, editor. *The social dimensions of science*. Notre Dame: University of Notre Dame Press; 1992. p. 130-57.

Recebido em 14/Jun/2013

Aprovado em 25/Jun/2013