

O autor responde

The author replies

El autor responde

<http://dx.doi.org/10.1590/0102-311XCO100913>*Kenneth Rochel de Camargo Jr.***Sigamos em frente?**

Escrevi o artigo que deu início a este debate sob o signo do desconforto, devido a um autoquestionamento sobre processos avaliativos dos quais participei. Escrevo a réplica a meus colegas com um sentimento de regozijo, dada a qualidade das respostas e daqueles que as escreveram, e agradeço a todos por terem aceito participar dessa discussão. A rigor talvez esta réplica nem fosse necessária, uma vez que o objetivo do texto inicial – convidar a um debate – já foi alcançado. Não posso, contudo, me furtar a discutir alguns aspectos gerais do que foi comentado.

Começo por agradecer a Rita Barradas Barata a correção de uma afirmação genérica – e por isso equivocada – que fiz, caracterizando indicadores bibliométricos como “médias”, quando de fato essa descrição não se aplica exatamente ao fator de impacto e menos ainda ao índice *h*. Tendo escrito artigo anterior, citado na primeira parte deste debate, não quis retomar toda a discussão crítica sobre os problemas decorrentes do uso de indicadores bibliométricos na avaliação da ciência, e fui excessivamente apressado na apresentação do sumário de alguns argumentos. Insisto, contudo, que uma das limitações para sua utilização como medida de impacto, importância ou qualidade de periódicos está na sua precária representação da distribuição geral de citações das respectivas revistas; como já alertava há mais de duas décadas Per Seglen¹, apenas 15% ou menos dos artigos publicados respondem por 80% ou mais do seu fator de impacto, e isso é pior ainda, por seu próprio desenho, no caso do índice *h*. Embora não seja de fato esse o maior problema dos indicadores bibliométricos, certamente é um componente relevante do quadro geral. Como Barata, considero extremamente problemática a avaliação de campos profundamente assimétricos, seja por qual critério for. Da mesma forma, a constituição de comitês julgadores que deem conta de tamanha variabilidade disciplinar e temática como encontramos na Saúde Coletiva é de fato altamente problemática. No entanto, não me parece que tais problemas sejam específicos de técnicas qualitativas para a avaliação, e se fa-

ção inevitavelmente presentes seja qual for a sistemática adotada.

Paulo Vaz introduz um aspecto de fundamental importância e – sintomaticamente – usualmente ignorado nas discussões sobre avaliação da ciência: o dos conflitos intracomunidade científica, mais claramente, das disputas de poder, ao mencionar o risco da adoção de técnicas de avaliação predominantemente qualitativas em “campos de conhecimento conflagrados”. Concordo que esse é um aspecto problemático, e mais uma vez torno a insistir que o é para qualquer forma de avaliação. Aqueles com mais poder em campos conflagrados terão sempre a possibilidade de criar normas que os favoreçam, mesmo utilizando critérios supostamente menos controversos por serem quantitativos, e seu sucesso nesse caso não sinalizaria, a meu ver, uma vantagem inerente da adoção de técnicas quantitativas, uma vez que não se estaria sendo necessariamente mais justo, mas, talvez, apenas menos controverso – o que paradoxalmente pode ser mais injusto ainda, ao restringir para campos não hegemônicos a possibilidade de manter a “conflagração” enquanto estiverem em posição desvantajosa.

Não gostaria, contudo, que a discussão se resumisse à falaciosa antinomia qualitativo/quantitativo, até porque, como lembra Hillegonda Maria Dutilh Novaes, dialeticamente há qualidade na quantidade e vice-versa, e a experiência internacional no campo, como ela mostra, aponta invariavelmente para um *mix* de abordagens. Não se trata aqui de uma recusa *a priori* a qualquer tipo de mensuração, mas sim de uma crítica à utilização continuada de métricas que já se mostraram sobejamente inadequadas para esse tipo específico de avaliação, e que em alguns casos em nosso país continuam sendo utilizadas quase que exclusivamente em tais processos. Nesse sentido, diria que concordo com Claudio José Struchiner (bem como com quase todos os comentaristas) sobre a necessidade de combinar diversos métodos em processos de avaliação, mas creio que deveríamos procurar outras métricas que não as em uso corrente. Também concordo com Struchiner que há uma lacuna conceitual sobre o que é qualidade científica, mas eu me atreveria a dizer que não é apenas no breve ensaio que abre esta discussão, mas em toda a produção sobre o tema. Acredito que esse tópico necessita de maior reflexão, e tenderia a questionar se tal definição, como o elusivo “critério de demarcação” que de uma vez por todas separaria a ciência da não-ciência, seria sequer possível de ser elaborada. Note-se, contudo, que a ausência de tal definição não impediu que se operassem ao longo da história da ciência várias formas de

avaliação, desde a elaboração de pareceres sobre artigos até decisões envolvendo a destinação de vultosos recursos para pesquisa. Diria mais, que talvez exatamente por uma impossibilidade de definição de uma vez por todas do que seja *qualidade* em ciência é que a revisão por pares segue sendo a espinha dorsal da ciência, por mais que o processo seja criticado. Indo além, ainda que reconhecendo a tendência ao autoengano e todos os problemas cognitivos que aponta, diria que não há como imaginar uma atividade de avaliação da ciência que proceda de modo radicalmente diferente da própria ciência, muito menos disciplinada no seu cotidiano do que gostariam os epistemólogos mais normativos, e que ainda assim, produz conhecimento.

A observação feita pelo professor Eduardo Massad, citada por Struchiner, de que diferentes processos avaliativos para professores em seu departamento chegaram a resultados semelhantes é interessante, mas acredito que teríamos resultados diversos em outros contextos. Tenho dúvidas se essa lógica, aplicada à produção de um pesquisador, teria o mesmo desfecho, isto é, se os produtos mais citados seriam invariavelmente aqueles que o pesquisador consideraria como sua produção mais relevante. No meu próprio caso, os trabalhos que consideraria mais relevantes na minha produção provavelmente não terão tal correspondência com os mais citados, a começar do fato de que parte do que considero mais relevante no que produzi foi publicada em livros e capítulos, que mal começam a ser contemplados por esse tipo de estatística.

Ainda com relação às observações feitas por Struchiner, gostaria de ressaltar que não proponho uma “volta ao passado”, com base em avaliações “subjetivas” e inevitavelmente personalistas. Por um lado, estratégias qualitativas não são necessariamente “subjetivas”; como aponta o próprio Struchiner, a suposta antinomia “qualitativo-subjetivo” *versus* “quantitativo-objetivo” carece de sentido e remete a uma querela ultrapassada na Saúde Coletiva. Por outro lado, as dificuldades e limitações com relação à composição de comitês de avaliação, insisto, estarão sempre presentes, não importa qual o processo adotado. Afinal, *quis custodiet ipsos custodes?*

O comentário de Luis Eugenio Portela Fernandes de Souza ressalta um ponto crucial, com o qual concordo plenamente: a adoção de novas sistemáticas – eu acrescentaria, quaisquer que sejam – não oferecem a garantia de avaliações justas e iguais. Como ele, acredito que resposta para esse tipo de problema não é metodológica, e sim política: quanto mais democráticos e transparentes forem a seleção e funcionamento de tais comitês, menor o espaço para que decisões dis-

cutíveis tenham lugar. Teria, entretanto, ressalvas às críticas que apresenta no final do texto. As críticas à utilização de indicadores bibliométricos como sinônimo de qualidade – e mesmo de impacto – são crescentes e contundentes, e o problema reside exatamente na falácia de que citação representa necessária e inequivocamente qualidade. Gostaria de chamar novamente a atenção para o artigo de Adler et al.², leitura obrigatória para esta discussão, ainda mais por ser a posição oficial de três associações internacionais das áreas de matemática e estatística, que afirmam, entre outras críticas, o seguinte: *“Aqueles que promovem as estatísticas de citações como medida predominante da qualidade da pesquisa não respondem à pergunta essencial: o que significam as citações? Eles juntam grandes quantidades de dados sobre o número de citações, processam os dados a fim de obter estatísticas e, em seguida afirmam que o processo de avaliação resultante é ‘objetivo’. No entanto, é a interpretação das estatísticas que leva à avaliação e a interpretação baseia-se no significado das citações, o que é bastante subjetivo”* (ênfases dos autores)² (p. 14). É nesse sentido que o título, ainda que provocativo, reflete a posição – que, repito, não é individual, veja-se a título de exemplo o manifesto DORA, citado no texto inicial – de que tais indicadores, tal como usados, criam de fato uma ficção contábil. Ainda me apoiando nesses autores, contesto a afirmação de que estaria desvalorizando a busca da objetividade; o problema não está no uso da noção de objetividade como ideal regulador, o que critico no texto é a ideia de que essa seria “alcançável” em absoluto. Ao acreditar que se alcançou a objetividade se deixa de buscá-la, esse é outro problema decorrente do uso acrítico da bibliometria.

Maurício L. Barreto aponta para a necessidade de uma reflexão mais profunda sobre relações de poder que levaram à estruturação de sistemas de avaliação tal qual existem, por um lado, e do desenvolvimento de uma ciência da avaliação científica, o que de certa forma se articula com a demanda feita por Struchiner de uma definição precisa do que seria qualidade na ciência. Vejo semelhança, mais uma vez, nesse anseio por uma espécie de metaciência normativa, ecos do projeto da ciência unificada (e mais uma vez do “critério de demarcação”) do positivismo lógico e de Popper; tendo em vista a direção tomada nas últimas décadas pela maior parte dos *science studies*, de abandono das grandes narrativas e foco no local, específico e contingente (ver, por exemplo, a crítica de Arthur Fine³), me pergunto se tal empreendimento seria até mesmo viável.

Colocaria ainda contra tais objeções o que parece ser uma tendência internacional, como

aponta Sonia Maria Ramos de Vasconcelos em seu comentário, de revisão de sistemáticas de avaliação em uso corrente, caracterizada de um modo geral pela redução da ênfase em, ou mesmo eliminação dos, indicadores bibliométricos, (re)introdução de técnicas qualitativas e valorização da revisão por pares. Como Vasconcelos indica no seu texto, esse processo implica a quebra de uma cultura estabelecida que, no limite, a meu ver impediria até que se concebessem formas alternativas de avaliação. Considero extremamente feliz a ponte que a autora faz com a discussão da ética e integridade na pesquisa, tanto no que diz respeito ao estabelecimento de formas mais justas de avaliação quanto no reconhecimento do impacto deletério que a lógica atual pode ter – e parece estar tendo – na produção e publicação da ciência, ao estimular uma produção mais voltada para colher recompensas do que para fazer ciência.

Gostaria de destacar no comentário de Nísia Trindade Lima os exemplos adicionais das repercussões negativas das formas institucionais de avaliação em curso, como a relegação de atividades antes consideradas nobres a um lugar subalterno ou a erosão do tempo de lazer dos pesquisadores, mas principalmente na sua concordância quanto à possibilidade de tais avaliações inibirem a inovação científica. Ressalto ainda no seu texto a recuperação do artigo de Carlos E. A. Coimbra Jr., publicado em 1999, e sua antecipação profética dos problemas relacionados à utilização de indicadores bibliométricos como métrica de qualidade científica.

Quanto a esse último, companheiro de sofrimento, não por acaso foi ao coração do que me motivou a escrever o texto inicial deste debate, a começar pelo próprio título de sua intervenção: com efeito, o que estamos avaliando? A questão da “zona cinza”, bem descrita em seu texto, escancara as dificuldades vividas por nós nos processos de avaliação, disparadoras da angústia a que me referi anteriormente. E compartilhamos da preocupação com a proliferação exponencial de autores de artigos, entre outros efeitos indesejados dos processos avaliativos e algo que compromete sua aplicação. É oportuna também a lembrança de que a alternativa que esbocei no texto inicial era na verdade parte do que propusemos, como comitê assessor do CNPq, como critério de julgamento de bolsistas de pesquisa em 2012, proposta infelizmente descartada por razões que até hoje desconheço. Curiosamente, na mesa de abertura do seminário de avaliação de bolsistas de produção científica que o CNPq promoveu no início de abril de 2013, tanto o Ministro da Ciência, Tecnologia e Inovação quanto o presidente do CNPq defenderam ⁴, em linhas

gerais, princípios semelhantes à proposta que elaboramos, e que reiterarei no texto inicial.

Esses são apenas alguns pontos dos diversos comentários que me suscitaram respostas. Não esgotam, de modo algum, a riqueza dos textos de meus colegas, muito menos o debate proposto. Para finalizar, gostaria de retomar alguns pontos do texto original, agora já informado pelas reações dos debatedores.

Em primeiro lugar, não me parece que nenhum dos comentaristas tenha feito uma defesa radical dos indicadores bibliométricos como medida de qualidade da produção científica. Gostaria de reiterar que a crítica que apresento não é fruto de idiosincrasia pessoal. É palpável o surgimento de uma mobilização da comunidade científica internacional contra a utilização de tais indicadores em avaliações, sendo a iniciativa DORA um exemplo mais visível e recente dessa resposta. Isso só torna ainda mais misteriosa a persistência de tais indicadores como pilares de processos avaliativos em nosso meio.

Em segundo lugar, a principal ressalva feita quanto à revalorização da revisão por pares nos processos avaliativos me parece ter sido em relação à possível parcialidade de avaliadores e o papel crítico que a composição de comitês passaria a ter por conta disso. Reitero que essa questão está posta em qualquer tipo de avaliação, e que a solução para ela é a máxima transparência possível em todo o processo. A necessária prudência para a implementação de mudanças não pode, no entanto, recair na paralisia da “covardia intelectual” tal como criticada por Reinach ⁵. É perfeitamente possível pactuar com a comunidade de pesquisa indicadores qualitativos objetivos para a avaliação; apenas a título de exemplo, acabo de participar num processo de consenso com base na metodologia Delphi promovida pela agência canadense de pesquisa em saúde (Canadian Institutes of Health Research/ Instituts de Recherche en Santé du Canada) para estabelecer critérios para a avaliação qualitativa de projetos de intervenção em saúde no nível populacional. Em qualquer caso, como lembra Vasconcelos, diversos países já adotam critérios qualitativos em suas avaliações da ciência, o que no mínimo demonstra a exequibilidade de sua adoção.

Por fim, gostaria de reiterar que, apesar de reconhecidamente não ter dado ênfase a esse aspecto no texto original, o que imagino como alternativa ao modelo atual não é a simples substituição do quantitativo pelo qualitativo, mas a integração de diferentes métodos e técnicas. A avaliação da ciência é complexa, e não terá soluções simples. A crítica que faço, repito, é a utilização tal como vem sendo feita de indicadores biblio-

métricos como necessários e suficientes para a avaliação, não sendo de modo algum uma recusa do quantitativo. Nosso procedimento no comitê assessor do CNPq, por exemplo, invariavelmente foi de tentar estabelecer um ranqueamento com base na produção dos candidatos à bolsa, e concentrar a avaliação naqueles com melhores posições. Talvez seja razoável supor que os melhores pesquisadores, como regra geral, ocupem o topo de tais ranqueamentos, mas regras têm exceções, e começar o processo com essa lógica impede que exceções sejam adequadamente contempladas. No que diz respeito à avaliação da produção no âmbito da Capes – apenas um indicador, decerto, mas com peso considerável e decisivo no conjunto da avaliação – a mudança do sistema anterior, com base em patamares, para o corrente, que transforma as diferentes publicações em pontos, foi certamente um avanço; produções importantes deixavam de ser consideradas por não estarem no patamar desejado. O sistema atual é certamente mais justo nesse aspecto, mas introduziu uma imprevisibilidade para professores, alunos e coordenadores dos programas, uma vez que a produção será também ranqueada entre os diversos programas, gerando a pressão por publicar cada vez mais e mais, em função do que ficou conhecido como “subir o sarrafo”, isto é, aumentar o nível de exigência a cada avaliação. No que diz respeito à quantidade de publicações, o céu passou a ser o limite.

Tanto no caso de pesquisadores individuais quanto de programas, talvez fosse mais razoável estabelecer patamares mínimos para a produção, a fim de garantir uma produção adequada e complementar essa avaliação com critérios qualitativos, como inovação temática e/ou metodológica, caracterização de linha(s) de pesquisa, coerência da produção com a linha de pesquisa e assim por diante. Como declarou Glaucius Oliva, presidente do CNPq, “*Não produzir nada ou pouco é sempre ruim e publicar muito também não é necessariamente bom. Temos que encontrar um equilíbrio*”⁴. Eu iria além da posição de Oliva e diria que os sistemas atuais de avaliação, que estimulam a publicação desenfreada de cada vez mais artigos, em especial, estão causando danos à produção científica, quer por estimular subterfúgios eticamente discutíveis, quer pela inibição da ousadia e da inventividade dos pesquisadores. Creio que a maior parte, senão a totalidade dos comentadores concordam que isso é um problema.

Não há, e possivelmente jamais haverá, sistema de avaliação perfeito, mas certamente qualquer sistema adotado é passível de aperfeiçoamento. O fundamental, a meu ver, é que utilizemos nossa imaginação criadora para en-

contrar novas formas de avaliação que permitam ao menos avançar em direção a um sistema que contemple de fato a “qualidade” da produção científica, seja lá como a definamos, e não termine por comprometê-la. Não podemos deixar que os riscos de efetuar mudanças nos paralisem num sistema insatisfatório.

Encerrando, deixo como desafio palavras de Fernando Pessoa, via Álvaro de Campos:

“Tens, como Hamlet, o pavor do desconhecido? Mas o que é conhecido? O que é que tu conheces, Para que chames desconhecido a qualquer coisa em especial?”

(Álvaro de Campos – *Se Te Queres*).

1. Seglen PO. The skewness of science. *J Am Soc Inf Sci* 1992; 43:628-38.
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. Fine A. Science made up: constructivist sociology of scientific knowledge. In: Gallison P, Stump DJ, editors. *The disunity of science: boundaries, contexts, and power*. Stanford: Stanford University Press; 1996. p. 231-54.
4. Ministro fala sobre a ciência brasileira em seminário de avaliação de bolsistas no CNPq. http://www.cnpq.br/web/guest/noticiasviews/-/journal_content/56_INSTANCE_a6MO/10157/955147 (acessado em 01/Ago/2013).
5. 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/2013).