



# NOVAS RÉGUAS PARA MEDIR A QUALIDADE

Universidades e agências de fomento buscam e adotam métricas mais abrangentes para examinar o desempenho de pesquisadores

Fabrcio Marques

**A** comunidade científica dos Países Baixos trava uma disputa que promete reverberar na forma como são avaliados a qualidade do trabalho e o desempenho de pesquisadores em outras partes do mundo. A controvérsia teve início em junho, quando a Universidade de Utrecht, a mais antiga e bem ranqueada instituição de ensino superior holandesa, anunciou uma reforma em suas regras de contratação e promoção, abolindo o uso de indicadores bibliométricos como o fator de impacto (FI) para mensurar a relevância da produção de seus docentes. Calculado pelo número de citações que um artigo recebe nas referências bibliográficas de outros artigos, o FI é adotado, por exemplo, para examinar o prestígio de periódicos científicos – a empresa Clarivate Analytics

atualiza anualmente o *Journal Citation Reports*, base de dados que estima o fator de impacto médio de mais de 10 mil revistas. O FI é considerado em diversas disciplinas como um bom sinalizador da repercussão que um artigo teve entre especialistas de sua área e se tornou matéria-prima para outros indicadores, como o índice h – que combina o número de artigos de um autor com a frequência de citações desses manuscritos.

No novo modelo proposto pela Universidade de Utrecht, os pesquisadores serão avaliados sem computar o número e a influência de seus *papers*, mas sim com base na qualidade do ensino, no compromisso de atuar em equipes e na disposição de compartilhar dados de pesquisa. Cada departamento deverá elaborar estratégias próprias para apreciar o desempenho de seus docentes, levando em conta o efeito na economia

e na sociedade e os princípios de “ciência aberta”, conjunto de práticas que promove a transparência e o trabalho em colaboração. “Temos uma forte convicção de que algo precisa mudar. E abandonar o fator de impacto é uma mudança necessária”, disse à revista *Nature* Paul Boselie, professor da Escola de Governança da universidade, responsável pela construção do novo sistema. Segundo ele, o uso indiscriminado do FI produziu consequências indesejadas, como a preocupação excessiva em publicar artigos e a busca de artifícios para anabolizar sua ressonância, deixando em segundo plano outros objetivos importantes da atividade científica.

A decisão gerou ondas de choque em outras instituições dos Países Baixos, que se materializaram em uma carta aberta assinada por 170 pesquisadores contrários à mudança e temerosos de que ela seja adotada por mais universidades – entre os signatários está o Nobel de Química de 2016 Bernard Feringa, da Universidade de Groningen. O contra-argumento é que, na ausência de indicadores objetivos, os processos de contratação e promoção passem a ser regidos por critérios eventualmente arbitrários. “O fator de impacto é uma métrica imperfeita, mas continua sendo útil”, disse à revista *Nature Index* um dos signatários da carta, o biólogo celular Raymond Poot, do Centro Médico da Universidade Erasmus, em Roterdã.

**N**a avaliação de Jacques Marcovitch, reitor da Universidade de São Paulo (USP) entre 1997 e 2001, o embate nos Países Baixos põe em evidência vantagens e limites das duas abordagens. “Indicadores bibliométricos são racionais e objetivos, mas eles sabidamente provocam distorções de comportamentos e são incapazes de captar dimensões como a qualidade do ensino na sala de aula”, afirma. Já a análise detalhada da contribuição científica e acadêmica de pesquisadores é mais trabalhosa e impõe desafios complexos. “Naturalmente, isso envolve um processo bem mais demorado e difícil de realizar”, afirma Marcovitch, que coordena um projeto financiado pela FAPESP voltado ao desenvolvimento de novas métricas para avaliar o desempenho científico, econômico e cultural das universidades públicas paulistas.

A disputa nos Países Baixos é simbólica porque estabelece uma ruptura com indicadores consagrados, cujo uso exagerado há tempos é criticado como reducionista. Nos últimos anos, uma série de manifestos propôs meios de fazer avaliações mais abrangentes e ganhou adeptos entre universidades de toda parte. O principal deles é a Declaração de São Francisco sobre Avaliação de Pesquisa (Dora), de 2012, endossada por mais de 20 mil pesquisadores e 2 mil instituições de

148 países, que recomenda abolir o uso isolado do fator de impacto de periódicos em avaliações relativas a financiamento, promoções e contratações. Outro documento de referência é um conjunto de diretrizes definidas em 2019 na 6ª Conferência Mundial de Integridade Científica, realizada em Hong Kong, para avaliar o desempenho dos pesquisadores de forma mais ampla e criar recompensas na carreira para os que adotarem práticas capazes de reforçar a integridade da ciência (ver Pesquisa FAPESP nº 303).

Instituições de vários países vêm reduzindo o peso de indicadores bibliométricos e ampliando o de parâmetros de caráter qualitativo, em busca do que se convencionou chamar de “métricas responsáveis”. A Universidade de Glasgow, no Reino Unido, incorporou recentemente o critério de “colegialidade” à avaliação de docentes: para serem promovidos ao topo da carreira, eles devem demonstrar que contribuíram com a trajetória de colegas e assistentes, compartilhando dados e produzindo artigos em coautoria. No sistema de avaliação das universidades do Reino Unido, utilizado para classificar as instituições e distribuir recursos, há tempos se conjugam alguns indicadores bibliométricos com a análise feita por pares de uma seleção dos trabalhos mais relevantes dos pesquisadores – o peso de cada modelo é ajustado a cada temporada de avaliação.

A estratégia da China também está em transformação. Em vez de valorizar o volume de estudos publicados, passou-se a exigir que os pesquisadores selecionem as melhores contribuições para serem analisadas por painéis de especialistas. Os chineses anunciaram a intenção de elaborar indicadores bibliométricos próprios, que contemplem o impacto regional de sua pesquisa.

Em novembro do ano passado, diversas iniciativas relacionadas a uma avaliação mais completa da pesquisa foram apresentadas em uma conferência virtual do Global Research Council, entidade criada em 2012 para estimular o intercâmbio de práticas de gestão entre agências de fomento. Em vários países, os pesquisadores estão sendo instados a fornecer uma narrativa estruturada

**Nos últimos anos, uma série de manifestos propôs meios de fazer avaliações mais amplas, ganhando adeptos entre as universidades**

sobre sua carreira, exprimindo sua contribuição individual, em vez de listar o volume de artigos e as citações que receberam. A Fundação Nacional de Ciência da Suíça está testando um currículo desse tipo, o SciCV, fácil de preencher e de atualizar. A Royal Society, do Reino Unido, desenvolveu um currículo dividido em quatro seções: geração de conhecimento, desenvolvimento de talentos, contribuição para a comunidade de pesquisa e contribuição para a sociedade.

**A** pesar das mudanças, os indicadores bibliométricos ainda são uma ferramenta muito presente na avaliação da ciência. Um estudo publicado em 2019 na revista *eLifeSciences* constatou que 40% das universidades de pesquisa nos Estados Unidos e Canadá mencionam fatores de impacto ou termos relacionados em documentos referentes à estabilidade, revisão e promoção de seus quadros. Um caso recente envolvendo a Universidade de Liverpool, no Reino Unido, é revelador das dificuldades de promover uma mudança de cultura. A instituição é uma das signatárias do manifesto Dora e se diz engajada na adoção de métricas responsáveis, como a análise por pares do trabalho de seus pesquisadores. Mas está sendo criticada por utilizar indicadores de

ordem financeira para escolher os 32 docentes de sua Faculdade de Saúde e Ciências da Vida que serão demitidos. Para manter os empregos, eles precisam demonstrar que conseguiram atrair financiamento para seus projetos em níveis semelhantes aos de um grupo de 24 universidades intensivas em pesquisa ao qual Liverpool pertence. Um editorial recente da revista *Nature* apontou o imbróglio de Liverpool como uma encruzilhada para o movimento deflagrado pelo manifesto Dora.

A busca e a adoção de métricas responsáveis também avançam no Brasil. Em um artigo divulgado recentemente nos *Anais da Academia Brasileira de Ciências*, um trio de bioquímicos lançou o manifesto “Avaliação científica responsável: Minimizando os índices, aumentando a qualidade”, que ressalta a importância da revisão por pares para identificar a contribuição de uma pesquisa. Uma das recomendações é criar mecanismos de recompensa para bons avaliadores, aqueles cientistas que têm conhecimento profundo em um tema e concentram suas sugestões na possibilidade de melhorar a qualidade do manuscrito ou de um projeto de pesquisa de colegas.

O documento propõe ainda que indicadores bibliométricos sejam utilizados com parcimônia, levando em conta as suas limitações. “O trabalho de pesquisadores atuando na fronteira do conhecimento não pode ser avaliado de forma quanti-

*Campus da Universidade de Utrecht: em vez do fator de impacto, docentes serão avaliados pela repercussão de seu trabalho no ensino, na economia e na sociedade*





tativa. A avaliação da qualidade só pode ser feita por pessoas experientes, na chamada revisão por pares”, afirma Alicia Kowaltowski, pesquisadora do Instituto de Química da USP e coordenadora de área da Diretoria Científica da FAPESP, que escreveu o manifesto em parceria com Ariel Silber, também da USP, e Marcus Oliveira, da Universidade Federal do Rio de Janeiro (UFRJ). O uso responsável de métricas, observa Kowaltowski, exige uma análise do contexto. “O número de citações varia de acordo com as áreas do conhecimento e é mediado por outros fatores – artigos de revisão, por exemplo, não trazem dados originais, mas costumam ser mais citados. O contexto é importante”, afirma.

O bioquímico Jorge Guimarães, que foi presidente da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) entre 2004 e 2015 e atualmente preside a Empresa Brasileira de Pesquisa e Inovação Industrial (Embrapii), vê com cautela a dicotomia entre indicadores de quantidade e qualidade. “Fala-se muito em mudar para componentes mais qualitativos, mas ninguém sabe muito bem quais deveriam ser usados”, afirma. Ele rejeita a ideia de que índices bibliométricos sejam meramente quantitativos. “O fator de impacto mede qualidade. Mostra que alguém leu seu artigo e o utilizou como referência, em geral por conta de sua contribuição.”

**N**os anos 2000, a Capes criou o Qualis, sistema de classificação de revistas científicas usado para avaliar os programas de pós-graduação do Brasil. O sistema está sendo revisto e costuma ser criticado por considerar a importância de um artigo não pelo número de citações que ele de fato recebeu, mas por um parâmetro indireto: o fator de impacto médio do periódico que o publicou. Essa estratégia é condenada em manifestos favoráveis a métricas responsáveis, mas Guimarães explica que ela teve uma razão de ser. “A avaliação dos programas leva em conta a produção científica gerada nos quatro anos anteriores. Nesse breve espaço de tempo, o número de citações de cada artigo é pequeno e não serviria como boa medida de avaliação”, afirmou. Ele observa que o peso atribuído a cada periódico era objeto de discussão aprofundada entre os representantes dos comitês de avaliação de cada área do conhecimento.

Guimarães sustenta que indicadores devem sempre ser interpretados com critério. Ele observa, por exemplo, que o impacto da ciência brasileira em agricultura tropical é baixo, uma vez que os resultados têm interesse regional. “Mas ninguém tem dúvida de sua importância econômica e social no Brasil. Isso precisa ser levado em conta na avaliação.” Já os pesquisadores brasileiros da

Universidade de Liverpool: críticas por utilizar indicadores financeiros para demitir pesquisadores

área de química, diz Guimarães, têm trabalhos de alta qualidade. “Mas há pouca transposição da pesquisa desenvolvida nas universidades para o setor industrial.”

**P**ara Jacques Marcovitch, um dos desafios é identificar os tipos de impacto que a universidade é capaz de gerar em diferentes áreas do conhecimento. “Há métricas que são adequadas para algumas disciplinas, mas não fazem nenhum sentido para outras. Se nas engenharias um indicador sobre patentes pode ser útil, na filosofia o reconhecimento dos pares é o principal objetivo a ser obtido”, explica. Já a sociedade espera das universidades um outro tipo de impacto. “Quer ver resultado todo ano, na forma de ingresso de novos alunos e na saída de profissionais bem formados, assim como espera extensão e pesquisa de qualidade. Na crise sanitária, essa pressão gerou um enorme estresse e as universidades se esmeraram em dar respostas à sociedade”, afirma.

A avaliação dos programas de pós-graduação da Capes está mudando para contemplar mais aspectos qualitativos. Alguns indicadores, como quantidade de professores e número de estudantes de mestrado e doutorado, perderam importância na avaliação quadrienal que se encerra neste ano e

não influenciarão mais na nota dos cursos. Serão considerados apenas como indicativo para manter o mínimo de pessoas atuando no programa. Já a publicação de artigos dos docentes e alunos vinculados a cada programa passará a ser analisada em três níveis diferentes e só o primeiro deles, o que mede a produção total, terá caráter quantitativo. Nos demais, contemplará análise por pares para avaliar um conjunto seletivo de trabalhos de cada docente, assim como os melhores itens da produção intelectual do programa – além da produção científica, também será possível escolher trabalhos técnicos e artísticos. Para o cientista da informação Rogério Mugnaini, da Escola de Comunicações e Artes da USP, as mudanças são interessantes. “A ideia de obrigar os docentes e a

coordenação do programa a destacar o que há de mais relevante na produção é boa e reduz o peso do volume de produção na avaliação do programa, desestimulando o produtivismo”, afirma. Segundo ele, ainda é cedo para avaliar quais serão os efeitos. “O ideal é que esses modelos fossem testados em um ciclo de avaliação, mas só implementados no seguinte, para ver se funcionam bem”, afirma. No futuro, a Capes planeja mudar profundamente seu sistema de avaliação, analisando os programas em cinco diferentes dimensões (ver Pesquisa FAPESP nº 286). De acordo com Mugnaini, a combinação ideal entre indicadores de quantidade e qualidade está por ser testada, mas ele não acredita que métricas sobre a produção venham a ser abandonadas. “Publicar trabalhos é parte essencial para a atividade científica e não me parece possível prescindir deles em um modelo de avaliação. Mas sem dúvida importa olhar além do mero produto, estimulando o desenvolvimento de projetos consistentes e duradouros e também a participação em redes de colaboração.”

A FAPESP tem aperfeiçoado o processo de avaliação de projetos para assegurar uma análise calçada em mérito e qualidade. A principal mudança envolve os termos usados nos formulários para apresentação de projetos – e servem para reforçar as expectativas da Fundação tanto para quem submete propostas quanto para quem as avalia. Em vez de indagar quais foram os artigos, livros ou patentes do proponente que tiveram mais impacto, o foco agora são os resultados científicos mais relevantes e a capacidade daquele projeto de multiplicar essa contribuição e ampliar seu alcance. “O objetivo é colocar na frente a qualidade do projeto e garantir que o que está sendo avaliado em primeiro lugar é a sua contribuição”, diz Cristóvão de Albuquerque, gerente de Colaboração em Pesquisa da FAPESP.

Há um esforço, também, no sentido de aperfeiçoar a revisão por pares. A Fundação fez um vídeo para orientar avaliadores que busca ajudá-los a produzir um despacho construtivo. “A ideia é que o parecer represente sempre uma contribuição à melhoria da proposta e, caso seja denegada, ela possa ser refeita e resubmetida adiante de forma mais consistente”, afirma Albuquerque. A FAPESP, observa ele, também permite que os proponentes incluam narrativas informando circunstâncias da vida relevantes para compreender sua contribuição. “Isso auxilia, por exemplo, pesquisadoras que foram mães e precisaram se afastar por um período”, conclui. ■

---

#### Projeto

Indicadores de desempenho nas universidades estaduais paulistas 2022 (nº 19/10963-7); Modalidade Pesquisa em Políticas Públicas; Pesquisador responsável Jacques Marcovitch (USP); Investimento R\$ 614.583,66.

**Quarenta por cento das universidades de pesquisa nos Estados Unidos e Canadá consideram o fator de impacto como critério de promoção de docentes**