

Avaliação da produção acadêmica

Assessment of academic production

Rita de Cássia Barradas Barata ¹

Resumo: Este artigo discute alguns aspectos da avaliação da produção científica, motivos, complexidade e razões para realizá-la. São apresentadas diferentes unidades de análise quais sejam as universidades, as faculdades, os grupos de pesquisa, os cursos de graduação ou pós-graduação, os pesquisadores ou docentes. Várias dimensões de avaliação dos produtos, tais como o volume da produção, a qualidade acadêmica, a relevância social, o impacto acadêmico e a inovação são brevemente discutidas. As abordagens quantitativas ou cientométricas e as abordagens qualitativas baseadas na revisão por pares são apresentadas com suas vantagens e desvantagens. Finalmente são apresentados alguns modelos de avaliação, combinando ou não esses enfoques.

Palavras-chave: avaliação da produção acadêmica; indicadores quantitativos; avaliação por pares.

Abstract: This essay discusses some aspects of the assessment of scientific production, arguments, complexity, and reasons for carrying it out. Are presented different units of analysis as universities, faculties, research groups, undergraduate or graduate courses, researchers, or professors. Several assessment dimensions, such as production volume, academic quality, social relevance, academic impact, and innovation are briefly discussed. Quantitative or scientometric approaches and qualitative approaches based on peer review are presented with their strengths and weaknesses. Finally, some assessment models are presented, combining or not these approaches.

Keywords: assessment of academic production; scientometric indicators; qualitative peer review.

¹ Faculdade de Ciências Médicas da Santa Casa de São Paulo.
São Paulo | São Paulo | Brasil.

Contato: rita.barradasbarata@gmail.com

Orcid: <https://orcid.org/0000-0002-7215-9788>

Recebido em: 18/06/2022

Revisado em: 06/12/2022

Aprovado em: 12/12/2022

DOI: <http://dx.doi.org/10.1590/S1414-40772022000300003>

Este é um artigo publicado em acesso aberto sob uma licença Creative Commons

<https://creativecommons.org/licenses/by-nc/4.0/>

1 Introdução

Nas últimas décadas do século XX observaram-se mudanças significativas nas relações entre ciência e sociedade com repercussões sobre a organização do campo científico. A participação da sociedade é cada vez mais presente na formulação das agendas de pesquisa, no controle da ética em pesquisa e na avaliação das atividades de investigação (PELLEGRINI FILHO, 2004).

O modo tradicional da pesquisa científica baseada em instituições de ensino e institutos de pesquisa; com agendas definidas por pesquisadores; enfoque disciplinar; transferência unilateral de conhecimentos para os usuários; avaliadas pelo mérito acadêmico; disseminada através de periódicos científicos; financiadas com recursos públicos e gerida de maneira centralizada, vai sendo substituído por um modo participativo de produção de ciência. O novo modelo caracteriza-se pelo trabalho em redes; agendas definidas em contexto de aplicação; pesquisas dirigidas para a solução de problemas; com enfoque transdisciplinar; intercâmbio permanente entre produtores e usuários; avaliação do mérito acadêmico e da relevância social; disseminada por múltiplos meios; financiada por diversas fontes públicas e privadas e gestão caracterizada por espaços de interação dos diversos atores (PELLEGRINI FILHO, 2004; BARRETO, 2013a).

As organizações de pesquisa são chamadas a prestar contas de suas atividades à sociedade. Esta prestação de contas deve responder a quatro princípios básicos: participação dos agentes no processo de tomada de decisão, avaliação interna e externa de desempenho, transparência sobre suas atividades e objetivos e estabelecimento de mecanismos de retroalimentação que incentivem comentários e críticas, internas e externas, necessárias à reorientação das estratégias institucionais. (WHITTY, 2008; CAMARGO JR., 2013; SOUZA, 2013).

Davyt e Velho (2000) afirmam que a avaliação além de ser parte integrante do processo de construção do conhecimento científico, é também o procedimento por meio do qual se definem os rumos do desenvolvimento científico e das instituições.

A avaliação da performance acadêmica tornou-se um fato, porém as métricas disponíveis não capturam a amplitude das atividades que dão suporte à produção e disseminação da produção e à formação de profissionais e acadêmicos. Os cientistas não gostam de ver a si mesmos e às suas instituições catalogadas e ranqueadas. Os argumentos contra os modelos de avaliação têm por base a compreensão de que o trabalho acadêmico é complexo e não pode ser avaliado de maneira simplista. (LANE, 2010).

Bastos (2013) assinala que a avaliação do trabalho acadêmico busca identificar o valor intrínseco da pesquisa da perspectiva dos pares, tornando a ciência e suas aplicações tecnológicas, bem como o impacto social de ambas, demonstráveis para os governantes e a sociedade, além de fornecer subsídios para a formulação de políticas em diferentes áreas da vida social.

Há várias razões para a realização de avaliações acadêmicas. A avaliação da performance é fonte de informação para as atividades de gestão, auxiliando a identificação das fortalezas e fraquezas que podem orientar as políticas institucionais e a alocação de recursos financeiros, materiais e humanos. A gestão acadêmica pode utilizar a avaliação para direcionar o desenvolvimento institucional, retribuição diferenciada aos docentes e pesquisadores segundo sua produtividade, apoio aos grupos que apresentam fragilidades. Finalmente, mecanismos de avaliação podem servir como instrumentos de monitoramento do trabalho executado. (STOKLAVA; TALASOVÁ; HOLECEK, 2011).

A avaliação da produção científica é um dos insumos importantes na formulação e implementação da política de ciência e tecnologia possibilitando o ajuste entre a capacidade instalada de pesquisa e as necessidades de desenvolvimento científico e tecnológico. A orientação do fomento também deveria estar baseada no maior conhecimento sobre a capacidade instalada das instituições e grupos de pesquisa e sobre a qualidade da sua produção.

Ainda em relação à orientação do fomento, as avaliações podem ajudar a formular estratégias para correção de assimetrias na distribuição da capacidade de pesquisa.

As avaliações das universidades e instituições de ensino têm sido usadas ainda para a construção de rankings de prestígio nacional e internacional e para conferir distinções no campo da política educacional. Este aspecto, embora controverso, acaba por ser incorporado pela sociedade e pelas próprias instituições como elementos do capital simbólico.

2 Unidades de Avaliação

Na elaboração dos rankings internacionais a unidade de avaliação é constituída pela Universidade como um todo. Nestes casos procura-se obter um conjunto de indicadores relativos ao número de alunos, número de docentes, relação alunos-docentes, número de artigos e citações etc.

Quando a avaliação é conduzida para propósitos administrativos a unidade de avaliação costuma ser constituída por unidades, faculdades ou institutos da Universidade, possibilitando distinguir o desempenho dos diferentes grupos. Pode ser interessante analisar cada departamento especificando ainda mais o foco da análise.

O Ministério da Educação realiza periodicamente a avaliação dos cursos de graduação utilizando nesse processo um conjunto de indicadores e relatórios de visitas realizadas por comissões de *experts*, além do resultado do Exame Nacional de Desempenho de Estudantes (ENADE).

Para programas de pós-graduação, a CAPES estabelece o formato utilizado na avaliação em todas as áreas, contemplando apenas três quesitos: Proposta do programa, Formação e Impacto na Sociedade reforçando a unidade de avaliação constituída pelo programa e deslocando a ênfase para a formação de docentes e pesquisadores.

O Diretório dos Grupos de Pesquisa do CNPq além de servir como base de registro de informações permite um conjunto de análises: titulação dos participantes, produção intelectual, atividades de orientação, linhas de pesquisa e setor de aplicação dos produtos.

Finalmente a avaliação da produção acadêmica pode ter como objeto cada pesquisador em particular seja para a obtenção de bolsas de produtividade em pesquisa, auxílio para participação em eventos científicos nacionais e internacionais, fomento para pesquisa, adicionais de produção, concursos de ingresso ou progressão na carreira acadêmica, credenciamento nos programas de pós-graduação.

Em cada uma dessas diferentes situações a avaliação precisa ser diferente e adotar indicadores apropriados ao que se pretende avaliar (NOVAES, 2013). Em todas elas, os artigos científicos recebem atenção porque a forma escrita é a que melhor documenta a produção científica e a torna acessível aos membros da comunidade de pesquisadores, seja no processo de construção dos consensos no interior dessas comunidades, seja no uso que os pesquisadores precisam fazer do saber acumulado para a produção de novas investigações. (BARRETO, 2013a; 2013b).

3 Dimensões

Considerando qualquer produto da produção acadêmica, há diferentes dimensões que poderiam ser analisadas: volume de produção, qualidade, impacto acadêmico, relevância social e inovação.

O **volume de produção** pode ser analisado em termos absolutos pelo número de produtos por unidade de tempo, ou em termos relativos através da taxa de produtos pelo número de pesquisadores. Os números absolutos não permitem a comparação entre grupos, programas, departamentos, universidades, dada a diversidade no número de pesquisadores. Os números absolutos serviriam apenas para comparação entre pesquisadores, mas mesmo nesse caso, seria necessário introduzir alguma correção pelo tempo de titulação.

Um indicador baseado no volume de produtos é facilmente obtido, e pode ser usado para um número grande de unidades de avaliação sem demandar tempo excessivo, desde que exista um registro desses produtos. No caso do Brasil há possibilidade de ter acesso à informação do número de artigos publicados pelos bolsistas de produtividade, pesquisadores dos grupos de pesquisa e docentes dos programas de pós-graduação a partir dos dados da plataforma Lattes, do diretório de grupos de pesquisa do CNPq e da Plataforma Sucupira da CAPES.

A **qualidade** desses produtos, definida por critérios de mérito acadêmico exige a análise por comissão de pares (*peer-review*) estando restrita a menor número de produtos selecionados, ou ao uso de critérios classificatórios genéricos tais como características dos veículos utilizados para divulgação, língua da publicação, ou outros. Nessa segunda abordagem haveria apenas a separação dos produtos em grupos sem que fosse possível estabelecer a hierarquia entre pesquisadores com base na qualidade do seu trabalho. Se o objetivo da avaliação for avaliar grupos, programas ou unidades pode ser suficiente apenas quantificar os produtos distribuídos nos grupos, porém se o objetivo for selecionar pesquisadores para bolsas de produtividade, ingresso ou progressão na carreira docente ou obtenção de financiamento para projetos, a análise de qualidade terá que ser mais precisa.

A vantagem da avaliação qualitativa do mérito acadêmico está na análise do produto em si, atendendo as demandas dos docentes que preferem ter seu trabalho avaliado pelo aspecto qualitativo ao invés da quantidade. Mas, há algumas desvantagens que devem ser consideradas: a análise qualitativa só pode ser aplicada a uma parte da produção e a escolha dos produtos a serem avaliados não é tarefa simples. Quem selecionaria os produtos? Que critérios seriam utilizados na avaliação? Como padronizar o julgamento dos pares?

A avaliação do **impacto acadêmico** dos artigos ou livros publicados pode ser feita através de indicadores de circulação e citação. Barreto (2013a) identifica as bases de indexação bibliográfica como os primeiros instrumentos concebidos para facilitar a circulação em uma época anterior ao desenvolvimento das ferramentas da informática. Atualmente, o potencial de difusão das bases indexadoras é incrivelmente maior visto que além de permitir a localização dos artigos muitas delas dão acesso ao texto completo facultando aos pesquisadores o acesso à produção científica sem a necessidade da intermediação de uma biblioteca física.

O desenvolvimento de métodos para recuperação das citações possibilitou a análise do impacto de cada artigo na construção de novos conhecimentos (BARRETO, 2013a; 2013b) enquanto o desenvolvimento de ferramentas de busca como o Google Scholar tornou a circulação ainda mais facilitada.

Há um conjunto de indicadores bibliométricos disponíveis para a avaliação do impacto acadêmico de artigos e dos periódicos onde eles estão publicados. Elas são medidas da possibilidade de um artigo ou um periódico influenciarem a produção do conhecimento em determinada área, baseada em seu potencial de circulação e em diferentes medidas de citações que refletem o uso desses conhecimentos por outros pesquisadores.

Outro critério que vem sendo utilizado na seleção de projetos é a **relevância social**. Na análise da produção acadêmica essa seria uma dimensão para considerar a aplicação potencial do conhecimento, seu uso para a solução de problemas concretos ou o acréscimo ao conhecimento. Valorização das aplicações práticas e do uso de conhecimentos científicos na fundamentação de documentos de políticas e na elaboração de normas e guias voltadas para a prática profissional complementariam o valor acadêmico.

Apesar do reconhecimento da importância de analisar esses aspectos não há instrumentos desenhados para analisar o potencial de aplicabilidade dos conhecimentos produzidos. A questão que se coloca é como avaliar essa incorporação sem que seja necessário realizar uma investigação específica para cada um dos produtos acadêmicos?

Os conhecimentos científicos resultantes do trabalho acadêmico podem ser utilizados de diferentes maneiras pelos gestores no processo de solução de problemas. Eles podem auxiliar a formulação dos problemas e indicar prováveis soluções; fornecer dados empíricos para realizar o diagnóstico do problema; auxiliar na modelagem de diferentes cursos de ação; servir como argumentos validados pela ciência em apoio a decisões tomadas na esfera política; ser usados de maneira instrumental ou ainda ser utilizados como capital simbólico na arena dos debates políticos (SOUZA; CONTRANDIOPOULUS, 2004).

Os políticos e gestores estão interessados na resposta a quatro questões relativas à efetividade (isto funciona?), resolubilidade (isto resolve o problema?), implementação (isto pode ser feito?) e custo-efetividade (isto pode economizar recursos?) Dificilmente os produtos das pesquisas acadêmicas podem responder prontamente a essas quatro perguntas. Geralmente há necessidade de uma etapa intermediária de “tradução” dos resultados das pesquisas para a prática e os tempos para que a incorporação se concretize são variáveis e indefinidos.

Atualmente vem se dando importância crescente à **inovação**, ou seja, a possibilidade dos produtos acadêmicos serem portadores de novidade. Do ponto de vista avaliativo este caráter inovador poderia ser identificado em diferentes etapas do processo de produção de conhecimento. A inovação poderia estar presente na formulação da pergunta de investigação ou manifestar-se na construção do método ou estar nos resultados em um novo saber ou em um novo modo de arranjar o conhecimento em modelos tecnológicos.

Avaliar esses diferentes aspectos da inovação combina as dificuldades já assinaladas para a avaliação do mérito acadêmico e da relevância social. Quem determina se uma questão de pesquisa, um método ou os resultados são inovadores? Uma apreciação teórico reflexiva ou o teste pragmático da prática? A divulgação de ideias inovadoras pode ser afetada pelo viés de publicação que torna mais difícil a aceitação de investigações que: questionam os paradigmas estabelecidos, as explicações aceitas, apresentam resultados paradoxais à luz do conhecimento dominante. Tal viés pode ocorrer tanto por uma posição conservadora dos editores quanto pela falta de parâmetros para uso dos avaliadores.

Portanto, são diversos os aspectos sob os quais cada produto acadêmico bibliográfico poderia ser analisado e cada um deles coloca desafios específicos para o desenvolvimento da avaliação.

4 Abordagens

Há duas vertentes principais na avaliação da produção acadêmica presentes isoladamente ou de maneira combinada nos diferentes modelos de avaliação acadêmica em uso: a abordagem cientométrica e a revisão por pares (CODINA, 2016).

As **abordagens cientométricas** envolvem questões relativas às bases de indexação bibliográficas e aos indicadores que podem ser construídos a partir das citações aos artigos ou periódicos.

Os dados da base de citações da *Web of Science* são os mais amplamente utilizados para análises cientométricas embora esta seja uma das bases bibliográficas mais restritas. A base *Scopus* é a mais ampla base de dados multidisciplinar existente no mercado. Ela inclui cerca de 25 mil periódicos revisados por pares. Há na base cerca de 15% de periódicos publicados em outras línguas que não o inglês embora haja um excesso relativo de periódicos norte-americanos, britânicos e holandeses. (MOYA-ANAGÓN *et al.*, 2007). Embora cada base apresente uma abrangência diferente, observa-se comparabilidade e estabilidade entre elas. A correlação entre o número de artigos e o número de citações recuperados em cada base é extremamente alta (ARCHAMBAULT; CAMPBELL; GINGRAS; LARIVIERE, 2009).

A escolha da base de dados pode influenciar os resultados dependendo da área de conhecimento. Barreto e colaboradores (2013c) estudaram a influência da fonte de dados no cálculo do índice-h de pesquisadores com bolsa de produtividade 1A do CNPq nas áreas de Imunologia, Medicina e Saúde Coletiva. As comparações feitas entre os índices obtidos da WoS, Scopus ou do Google Acadêmico mostraram diferenças não significantes para os

pesquisadores de Imunologia e diferenças significantes para pesquisadores de Medicina e Saúde Coletiva.

Os indicadores bibliométricos disponíveis para a avaliação de periódicos científicos são o fator de impacto ou citações por documento, o índice-h, o SJR, a meia vida das citações, o número médio de citações por artigo, a proporção de autocitações. A análise da performance da produção científica de um país ou em uma área específica do conhecimento pode incluir a proporção de artigos com cooperação internacional, a proporção de citações recebidas de periódicos do mesmo país, a proporção da publicação do país em relação à produção mundial, a proporção da produção daquela área no país em relação à proporção da publicação mundial naquela área, a proporção da produção do país divulgada em periódicos do próprio país ou de outros países, a proporção de artigos publicados na língua nativa ou em outras línguas, entre outros. Para o estudo do desempenho dos pesquisadores podem ser utilizados o número de citações dos artigos publicados ao longo da vida ou em períodos determinados, o índice-h, a proporção de autocitações, a proporção de artigos publicados em inglês, a distribuição dos artigos segundo classificação das revistas de circulação nacional ou internacional ou segundo o impacto do veículo utilizado.

O fator de impacto é visto incorretamente como um indicador de qualidade do periódico ignorando as forças que movem o processo de citação. Artigos podem ser citados meramente por aspectos técnicos, metodológicos ou tópicos; porque é científica ou culturalmente correto citar; porque os autores citados têm poder na comunidade científica; ou mesmo incorretamente por má compreensão do autor. (PORTA *et al.*, 2003; DECLARAÇÃO DE SÃO FRANCISCO ..., 2017).

O fator de impacto, assim como a contagem do número de citações de um artigo, sofre a influência de uma série de fatores tais como o número de periódicos indexados na base; a língua de publicação do artigo; o tipo de artigo; campo do conhecimento e temática; meia vida dos artigos em cada área do conhecimento; número médio de autores por artigo. (RUANO-RAVINA; ÁLVAREZ-DARDET, 2012).

Na medida em que o fator de impacto teve seu uso ampliado em diferentes contextos avaliativos, várias formas de manipulação dos resultados passaram a ser observadas, denotando má conduta editorial ou mau uso do indicador. Dentre as práticas mais comuns de manipulação do fator de impacto pelos editores dos periódicos estão a indução à citação de artigos previamente publicados pelo periódico; a redução artificial do denominador inflando a publicação de material considerado como “não citável”; a ampliação da publicação de revisões; inflacionar as citações através de editoriais e comentários de leitores; rejeição de estudos com

resultados negativos; rejeição de estudos confirmatórios; favorecimento a artigos produzidos por grandes grupos de pesquisa e com grande número de autores; atração de artigos de autores renomados; favorecimento à publicação de artigos que tratam de tópicos “quentes”. (FALAGAS; ALEXIOU, 2008).

Como as bases indexadoras passaram a calcular o fator de impacto após a eliminação da autocitação, uma nova modalidade de manipulação foi detectada. Na tentativa de ampliar as citações sem incorrer em um aumento considerado inaceitável (acima de 20%) da autocitação, alguns editores apelaram para um mecanismo de “troca de citações” realizadas através de artigos de revisão que possibilitam um maior número de referências e consequentemente de citações. Estes artigos são intencionalmente escritos, em geral com autoria dos próprios editores e outros pesquisadores, para permitir a troca de citações entre periódicos.

A Declaração de São Francisco (2017) recomenda não usar as métricas baseadas no desempenho das revistas como medida substitutiva da qualidade dos artigos individuais. A proposta é considerar um conjunto de medidas de impacto qualitativas e quantitativas da própria pesquisa, ou seja, de cada artigo individualmente.

Uma métrica melhor para avaliar a produção individual é o índice-h proposto por Jorge Hirsch em 2005. O índice-h combina produtividade e impacto, os dados necessários são fáceis de obter, não é sensível a valores extremos e é difícil de falsificar. (BATISTA; CAMPITELI; KINOCHI; MARTINEZ, 2006). O índice-h é uma medida tempo dependente, isto é, periódicos mais antigos e pesquisadores titulados a mais tempo tendem a apresentar valores maiores. Ele é sensível ao campo de conhecimento em função das dimensões das comunidades científicas e de suas práticas de produção de conhecimentos com maior ênfase na produção individual ou em grupos ou redes de pesquisadores. Também a língua de publicação pode influenciar o índice e a base de dados utilizada em seu cálculo.

De modo geral todos os autores alertam para a necessidade de combinar diferentes medidas em qualquer tipo de avaliação e assinalam a importância de que os avaliadores conheçam bem as medidas para que possam usá-las de maneira apropriada evitando conclusões não autorizadas.

A outra abordagem presente nas avaliações, de maneira isolada ou em combinação com as medidas bibliométricas é a **abordagem baseada na revisão por pares**. Historicamente, a revisão por pares surgiu para a avaliação da pesquisa já realizada, com vistas à publicação. O sistema de consulta aos pares previamente à aceitação para publicação faz parte do processo de construção de verdade científicas através do consenso entre os pares. Os pares julgam de acordo com os paradigmas e as normas estabelecidas pela comunidade disciplinar (DAVYT; VELHO,

2000). Do julgamento de pesquisas realizadas, o processo expandiu-se para o julgamento das propostas ou projetos e para outras avaliações de desempenho na comunidade científica. (CODINA, 2016).

O processo de revisão por pares tem sido objeto de muitas críticas relacionadas ao conservadorismo, baixa eficiência, autoritarismo, tendenciosidade, confiabilidade e problemas éticos. O sistema de *peer-review* é visto como conservador por resistir a ideias novas, conferir grande poder à burocracia das agências de fomento ou aos órgãos encarregados das avaliações, e pelo caráter normativo das avaliações que não são capazes de conter a diversidade e o dinamismo do trabalho científico. O processo é considerado pouco eficiente tendo em vista o tempo e o esforço demandado dos avaliadores. O fato dos avaliadores serem protegidos pelo sigilo somado à impossibilidade de recurso das decisões é visto como fonte de autoritarismo. A tendenciosidade observada na composição das comissões, geralmente formadas por pesquisadores ou instituições de prestígio, é apontada por jovens cientistas ou instituições sem tradição, como fonte de descontentamento. A confiabilidade é baixa visto que o desacordo entre avaliadores é muito frequente, tornando o sistema vulnerável e pouco estável. As questões éticas e os conflitos de interesse são frequentemente invocados em um sistema que favorece a competição ao invés de fortalecer a cooperação e a colaboração entre cientistas. (DAVYT; VELHO, 2000; CODINA, 2016).

A primeira dificuldade da avaliação por pares é a própria obtenção dos avaliadores. Tem sido cada vez mais difícil tanto para os periódicos científicos quanto para as agências de fomento, conseguir um número mínimo de avaliadores ou compor comissões de julgamento. Tite e Schroter (2007) estudaram os motivos de aceitação ou recusa dos avaliadores e quais seriam os incentivos que poderiam encorajá-los a aceitar a tarefa. Dentre os motivos apontados para a aceitação do convite foram mencionados: contribuição para a área de conhecimento do avaliador, relevância do tópico para o trabalho do avaliador, oportunidade de aprender algo novo, senso de responsabilidade profissional e reputação do periódico.

As razões apontadas pelos entrevistados para declinar do convite foram: falta de tempo para fazer a avaliação solicitada, ter muitas avaliações de outros periódicos para serem feitas, e prazos muito curtos para a análise. A maioria dos entrevistados considerou que os incentivos financeiros não seriam efetivos face às limitações de tempo. Os incentivos que foram considerados encorajadores foram acesso livre aos conteúdos do periódico, agradecimento no site do periódico, feedback sobre o julgamento final e sobre a qualidade da avaliação e convite para compor o grupo de editores (TITE; SCHROTER, 2007). Embora a pesquisa tenha sido dirigida a revisores de artigos submetidos a periódicos, os motivos de aceitação e recusa, e a

posição relativa aos incentivos podem ser consideradas pertinentes a outras formas de avaliação.

O segundo aspecto importante relativo ao processo de revisão por pares é a qualidade da revisão. Fora do âmbito editorial a questão da qualidade da avaliação torna-se em geral ainda mais complexa uma vez que as comissões de avaliação de projetos, programas, departamentos ou instituições são formadas por um número relativamente restrito de avaliadores com a tarefa de analisar situações bastante distintas; com visões diferentes sobre ciência; pertencendo a comunidades epistêmicas diversas; e, não compartilhando dos mesmos critérios para a avaliação de desempenho.

O terceiro ponto polêmico se refere à escolha dos avaliadores. Os conflitos de interesse aparecem como motivos de grande insatisfação nos processos avaliativos. Esses interesses podem ser individuais, institucionais, financeiros, acadêmicos ou pessoais. Os pesquisadores, consideram que diferentes tipos de conflitos de interesse podem afetar os resultados de processos avaliativos. Dentre eles são mencionados os conflitos disciplinares, decorrentes da competição entre áreas ou escolas de pensamento; rivalidade ou nepotismo; competição entre instituições. O principal mecanismo de regulação contra conflitos de interesse consiste em utilizar mais de um avaliador para os mesmos projetos, programas ou instituições e se possível mesclar diferentes origens. (ABDOUL *et al.*, 2012; CODINA, 2016).

Algumas sugestões para melhorar o processo de revisão por pares incluem o anonimato dos avaliados quando possível, seleção apropriada de avaliadores considerando as especialidades ou tópicos de pesquisa, oportunidade de recurso das decisões dos avaliadores, maior transparência do processo, interação entre avaliados e avaliadores, explicitação dos conflitos de interesse, treinamento dos avaliadores e estabelecimento de critérios mais uniformes. (ABDOUL *et al.*, 2012; CODINA, 2016).

Portanto, nenhuma das abordagens adotadas para a avaliação da produção acadêmica é completamente satisfatória e isenta de problemas. Cada uma possui vantagens e desvantagens que precisam ser consideradas na proposição de modelos avaliativos.

Abbott e colaboradores (2010) afirmam que nenhuma carreira pode ser resumida a um número. Os cientistas gastam horas incontáveis concebendo experimentos, orientando estudantes, escrevendo e revisando projetos e artigos, ensinando, preparando e organizando reuniões científicas, participando em colaborações, trocando ideias com colegas, atuando em corpos editoriais e comitês, e nenhuma dessas atividades é facilmente quantificável.

A revista *Nature*, em 2010, realizou um inquérito entre leitores e entrevistas com reitores, chefes de departamentos e administradores de 30 instituições em vários países do

mundo para estudar a percepção dos pesquisadores e dos gestores quanto ao uso de métricas nos processos de avaliação acadêmica. Cerca de 2/3 dos pesquisadores acreditavam que as métricas estavam sendo usadas nas decisões sobre contratação, promoção e revisão do desempenho acadêmico. Cerca de 60% estavam infelizes com esses processos. Já nas entrevistas com os gestores eles disseram que as métricas não importavam tanto quanto os pesquisadores pensavam e muitos afirmaram que davam maior importância às cartas de recomendação, currículo e entrevista, além da opinião da comissão de seleção sobre os trabalhos publicados (ABBOTT *et al.*, 2010).

Apesar da insatisfação com o modo como as métricas estavam sendo usadas alguns pesquisadores consideravam que o uso de métricas era bem-vindo dada a transparência e objetividade. Principalmente os jovens cientistas se sentiam mais seguros com o uso dessas medidas pois segundo eles as métricas poderiam assegurar que não se perpetuasse nas instituições uma rede de pesquisadores mais velhos na qual as relações pessoais fossem mais valorizadas que o desempenho. Surpreendentemente, se os pesquisadores desejavam mudanças elas não iam necessariamente na direção contrária ao uso das métricas. O inquérito solicitou que os entrevistados indicassem critérios que considerassem os mais adequados para avaliar pesquisadores. Os quatro critérios mais indicados foram: publicação em periódicos de alto impacto; obtenção de financiamentos a projetos; orientação de estudantes e número de citações em publicações. (ABBOTT *et al.*, 2010).

Assim, o desafio da avaliação parece estar não exatamente na redução das métricas, mas sim em um uso mais claro, consistente e transparente. Além disso, há um consenso sobre a necessidade de desenvolver indicadores melhores.

5 Modelos de Avaliação

A avaliação da produção acadêmica tem evoluído dos processos mais simples para procedimentos altamente complexos de estabelecimento de padrões e modelos envolvendo grande número de critérios qualitativos e padrões de performance e sistemas de contabilidade para contar praticamente tudo. Há basicamente três tipos de modelos de avaliação em uso em países europeus e da América do Norte: o modelo de performance em larga escala utilizado para alocação de recursos entre universidades e institutos de pesquisa; o modelo de alocação de recursos em blocos para grandes grupos; o modelo baseado em algoritmos usando número de estudantes e indicadores bibliométricos. (CORYN; HATTIE; SCRIVEN; HARTMANN, 2007).

O *Research Assessment Exercise (RAE)* foi utilizado pelo governo britânico, desde a década de 80, para estabelecer um ranking entre as universidades britânicas e orientar a alocação de recursos para pesquisa. O processo estava amplamente baseado em revisão por pares. Cada grupo ou departamento avaliado era classificado em uma escala de 1 a 5. (CORYN *et al.*, 2007). Em 2014 o RAE foi substituído pelo *Research Excellence Framework (REF)* que combina a revisão por pares com métricas tais como os índices de citação e uma avaliação da relevância econômica e social da pesquisa.

Na percepção da comunidade acadêmica os modelos de avaliação são insatisfatórios e não são capazes de avaliar a complexidade do trabalho acadêmico. Nóvoa (2017), ex-reitor da Universidade de Lisboa, destaca quatro questões fundamentais para compreender o mal-estar que vem afetando a vida acadêmica. A primeira questão refere-se à busca pela excelência como motora dos planos estratégicos das instituições de ensino que se traduz em incentivos ao produtivismo. A segunda questão diz respeito ao empreendedorismo que transforma as universidades em empresas introduzindo práticas de gestão estranhas à cultura acadêmica e contrariam a prática e o processo de trabalho da ciência. A terceira questão, remete aos objetivos de formação direcionados para a preparação de profissionais para os futuros empregos abdicando da missão educacional e cultural mais ampla. Finalmente, o autor faz uma crítica à europeização, ou seja, a tendência a apagar as diferenças culturais e as tradições dos países em prol de uma única visão europeia.

No Brasil há três modelos principais de avaliação da produção acadêmica regularmente em uso: a avaliação dos cursos de graduação realizada pelo MEC, a avaliação dos programas de pós-graduação realizada pela CAPES e a avaliação dos pesquisadores individuais realizada pelo CNPq para a concessão de bolsas de produtividade em pesquisa. O processo mais transparente é o da avaliação dos programas de pós-graduação para a qual os documentos de área, a ficha de avaliação e os indicadores utilizados são amplamente divulgados. Na avaliação dos cursos de graduação são divulgadas apenas as notas e apenas as instituições avaliadas têm acesso aos relatórios mais detalhados. O procedimento menos transparente é o dos comitês assessores do CNPq, pois nem sempre as deliberações dos comitês são acatadas pela direção da agência.

É necessário desenvolver novas ferramentas de avaliação que combinem indicadores quantitativos e qualitativos e modelos que incluam de forma mais equilibrada os diferentes aspectos do trabalho acadêmico. Também é importante ter claro que o processo avaliativo não precisa contabilizar todas as atividades docentes e discentes, mas identificar aquelas atividades nucleares que possam ser usadas como proxy para o conjunto. Entretanto, se a avaliação não

vier associada a algum tipo de reconhecimento dos diferentes perfis e das diferentes performances dos docentes, ela só servirá para estimular comportamentos competitivos e gerar insatisfação.

Referências

- ABBOTT, A. *et al.* Do metrics matter? **Nature**, Berlin, v. 465, p. 860-862, June, 2010. DOI 10.1038/465860a. Disponível em: www.pubmed.ncbi.nlm.nih.gov. Acesso em: 10 dez. 2022.
- ABDOUL, H. *et al.* Non-financial conflicts of interest in academic Grant evaluation: a qualitative study of multiple stakeholders in France. **PloS One**, San Francisco v. 7, n. 4, e35247, 2012. Disponível em: <https://doi.org/10.1371/journal.pone.0035247>. Acesso em: 10 dez. 2022.
- ARCHAMBAULT, E.; CAMPBELL, D.; GINGRAS, Y.; LARIVIERE, V. Comparing bibliometric statistics obtained from the Web of Science and Scopus. **Journal of the American Society for information science and technology**, Hoboken, v. 60, n. 7, p. 1320-1326, 2009. Disponível em: <https://doi.org/10.1002/asi.21062>. Acesso em: 10 dez. 2022.
- BARRETO, M. L. O desafio de avaliar o impacto das ciências para além da bibliometria. **Revista de Saúde Pública**, São Paulo, v. 47, n. 4, p. 834-837, 2013a. Disponível em: <https://doi.org/10.1590/S0034-8910.2013047005073>. Acesso em: 10 dez. 2022.
- BARRETO, M. L. Como avaliar as ciências com uma deficiente ciência da avaliação científica? **Cadernos de Saúde Pública**, Rio de Janeiro, v. 29, n. 9, p. 1719-1721, 2013b. Disponível em: <http://dx.doi.org/10.1590/0102-311XCO060913>. Acesso em: 10 dez. 2022.
- BARRETO, M. L. *et al.* Diferenças entre as medidas do índice h geradas em distintas fontes bibliográficas e engenho de busca. **Revista de Saúde Pública**, São Paulo, v. 47, n. 2, p. 231-238, 2013c. Disponível em: <https://doi.org/10.1590/S0034-89102013000100008>. Acesso em: 10 dez. 2022.
- BASTOS, F. I. P. M. Challenges and perspectives of academic evaluation. **Revista de Saúde Pública**, São Paulo, v. 47, n. 4, p. 829-833, 2013. Disponível em: <https://doi.org/10.1590/S0034-8910.2013047004988>. Acesso em: 10 dez. 2022.
- BATISTA, P. D.; CAMPITELI, M. G.; KINOUCI, O.; MARTINEZ, A. S. Is it possible to compare researchers with different scientific interests? **Scientometrics**, Budapest, v. 68, n. 1, p. 179-189, 2006. Disponível em: <https://doi.org/10.1007/s11192-006-0090-4>. Acesso em: 10 dez. 2022.
- CAMARGO JR., K. R. Produção científica: avaliação da qualidade ou ficção contábil? **Cadernos de Saúde Pública**, Rio de Janeiro, v. 29, n. 9, p. 1707-1710, 2013. Disponível em: <https://doi.org/10.1590/0102-311X00115413>. Acesso em: 10 dez. 2022.
- CODINA, L. Evaluación de la ciencia: tan necesaria como problemática. **El profesional de la información**, España, v. 25, n. 5, p. 715-719, 2016. Disponível em: <https://doi.org/10.3145/epi.2016.sep.01>. Acesso em: 10 dez. 2022.
- CORYN, C. L. S.; HATTIE, J. A.; SCRIVEN, M.; HARTMANN, D. J. Models and mechanisms for evaluating government-funded research: an international comparison. **American Journal of Evaluation**, New York, v. 28, n. 4, p. 437-457, 2007. Disponível em: <https://doi.org/10.1177/>. Acesso em: 10 dez. 2022.

DAVYT, A.; VELHO, L. A avaliação da ciência e a revisão por pares: passado e presente. Como será o futuro? **História, Ciências, Saúde-Manguinhos**, Rio de Janeiro, v. 7, n. 1, p. 93-116, 2000. Disponível em: <https://doi.org/10.1590/S0104-59702000000200005>. Acesso em: 10 dez. 2022.

DECLARAÇÃO DE SÃO FRANCISCO sobre a avaliação da pesquisa: introduzindo a ciência na avaliação da pesquisa. **Revista da ADUSP**, São Paulo, v. 60, p. 94-97, maio, 2017. Disponível em: www.adusp.org.br. Acesso em: 10 dez. 2022.

FALAGAS, M. E.; ALEXIOU, V. G. The top-ten in journal impact factor manipulation. **Archivum Immunologiae et Therapia Experimentalis**, Wroclaw, v. 56, p. 223-226, 2008. Disponível em: <https://doi.org/10.1007/s00005-008-0024-5>. Acesso em: 10 dez. 2022.

LANE, J. Let's make science metrics more scientific. **Nature**, Berlin, v. 464, p. 488-489, mar. 2010. DOI: [10.1038/464488a](https://doi.org/10.1038/464488a). Disponível em: www.ncbi.nlm.nih.gov. Acesso em: 10 dez. 2022.

MOYA-ANAGÓN, F. *et.al.* Coverage analysis of Scopus: a journal metric approach. **Scientometrics**, Budapest, v. 73, n. 1, p. 53-78, 2007. Disponível em: <http://dx.doi.org/10.1007/s11192-007-1681-4>. Acesso em: 10 dez. 2022.

NOVAES, H. M. D. Muitas dúvidas, poucas certezas, enquanto isso... **Cadernos de Saúde Pública**, Rio de Janeiro, v. 29, n. 9, p. 1715-1716, 2013. Disponível em: <http://dx.doi.org/10.1590/0102-311XCO030913>. Acesso em: 10 dez. 2022.

NÓVOA, A. Em busca da liberdade nas universidades, é tempo de dizer não. Publicado originalmente na Revista Lusófoma de Educação em 2014 e republicado pela **Revista da ADUSP**, São Paulo, v. 60, p. 46-53, maio, 2017. Disponível em: www.adusp.org.br. Acesso em: 10 dez. 2022.

PELLEGRINI FILHO, A. Pesquisa em saúde, política de saúde e equidade na América Latina. **Ciência & Saúde Coletiva**, Rio de Janeiro, v. 9, n. 2, p. 339-350, 2004. Disponível em: <https://doi.org/10.1590/S1413-81232004000200011>. Acesso em: 10 dez. 2022.

PORTA, M. *et.al.* Mixing journal, article, and author citations, and other pitfalls in the bibliographic impact factor. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 19, n. 6, p. 1847-1862, 2003. Disponível em: <https://doi.org/10.1590/S0102-311X2003000600030>. Acesso em: 10 dez. 2022.

RUANO-RAVINA, A.; ALVAREZ-DARDET, C. Evidence-based editing: factors influencing the number of citations in a national journal. **Annals of Epidemiology**, Amsterdam, v. 22, n. 9, p. 649-653, 2012. Disponível em: <https://doi.org/10.1016/j.annepidem.2012.06.104>. Acesso em: 10 dez. 2022.

SOUZA, L. E. P. F.; CONTRANDIOPOULOS, A-P. O uso de pesquisas na formulação de políticas de saúde: obstáculos e estratégias. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 20, n. 2, p. 546-554, 2004. Disponível em: <https://doi.org/10.1590/S0102-311X2004000200023>. Acesso em: 10 dez. 2022.

SOUZA, L. E. P. F. O desafio da avaliação da produção científica. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 29, n. 9, p. 1717-1719, 2013. Disponível em: <https://doi.org/10.1590/0102-311XCO050913>. Acesso em: 10 dez. 2022.

STOKLASA, J.; TALASOVÁ, J.; HOLECEK, P. Academic staff performance evaluation-variants of models. **Acta Polytechnica Hungarica**, Budapest, v. 8, n. 3, p. 91-111, 2011. Disponível em: www.acta.uni.obuda.hu. Acesso em: 10 dez. 2022.

TITE, L.; SCHROTER, S. Why do peer reviewers decline to review? A survey. **Journal of Epidemiology and Community Health**, Oxford, v. 61, p. 9-12, 2007. Disponível em: <http://www.ncbi.nlm.nih.gov/pmc/articles/pmc2465595/>. Acesso em: 10 dez. 2022.

WHITTY, B. **Accountability principles for research organizations**. One World Trust. Briefing paper 113. London, 2008. Disponível em: www.oneworldtrust.org. Acesso em: 10 dez. 2022.