

ESTRATÉGIAS DE AVALIAÇÃO DA PRODUÇÃO CIENTÍFICA: REFLEXÕES SOBRE SUAS LIMITAÇÕES E COMPLEMENTARIDADES

Rita Barradas Barata

 <http://lattes.cnpq.br/2280062846829854> –  <https://orcid.org/0000-0002-7215-9788>
rita.barradasbarata@gmail.com

Faculdade de Ciências Médicas Santa Casa de São Paulo
São Paulo, SP, Brasil

RESUMO

O ensaio foi construído a partir da análise da literatura pertinente e da experiência da autora em processos avaliativos da política científica, avaliação de programas de pós-graduação *estricto sensu* e na atividade editorial em periódicos científicos publicados no país e no exterior. A avaliação da produção científica através dos artigos científicos tem duas grandes vertentes principais: a avaliação dos artigos em si e a avaliação dos periódicos nos quais eles são publicados. Na avaliação dos artigos em si, podem ser usadas fundamentalmente três estratégias distintas e complementares: a classificação segundo características formais do artigo, a contagem das citações recebidas, e a avaliação qualitativa baseada nos procedimentos de *peer review* pós-publicação. Na avaliação dos periódicos como indicadores indiretos da qualidade da produção, utilizam-se diferentes indicadores bibliométricos, seja como fator de ponderação dos artigos, seja como critérios para construir classificações como o Qualis periódicos da CAPES. Como em toda questão complexa, não pode haver uma única forma de realizar a avaliação e, o uso de diferentes estratégias pode contrabalançar limitações existentes em cada uma delas, dando uma perspectiva mais abrangente e multifacetada para os avaliadores.

Palavras-chave: Artigos científicos. Periódicos científicos. Avaliação da produção científica.

STRATEGIES FOR EVALUATING SCIENTIFIC PRODUCTION: REFLECTIONS ON ITS LIMITATION AND COMPLEMENTARITIES

ABSTRACT

The essay was built from the analysis of the relevant literature and also from the author's experience in scientific policy evaluation processes, evaluation of strict sense graduate programs and in editorial activity in scientific journals published in the country and abroad. The evaluation of scientific production through scientific articles has two main aspects: the evaluation of the articles themselves and the evaluation of the journals in which they are published. In the evaluation of the articles themselves, three different and complementary strategies can be fundamentally used: classification according to the formal characteristics of the article, counting of citations received, and qualitative evaluation based on post-publication peer review procedures. In evaluating journals as indirect indicators of production quality, different bibliometric indicators are used, either as a weighting factor for articles numbers, or as criteria for building classifications such as Qualis journals from CAPES. As in any complex issue, there cannot be a single way of carrying out the evaluation, and the use of different strategies can counterbalance existing limitations in each of them, giving a more comprehensive and multifaceted perspective for the evaluators.

Keywords: Scientific articles. Scientific journals. Scientific production evaluation.

DOI <http://dx.doi.org/10.1590/1981-5344/42242>

Recebido em: 26/12/2022.
Aceito em: 27/10/2023.

1 INTRODUÇÃO

A avaliação da produção científica é um insumo importante 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. Do mesmo modo, este conhecimento poderia informar as estratégias de fomento e a formação de pesquisadores. As avaliações podem ajudar a formular estratégias para correção de assimetrias na distribuição da capacidade de pesquisa, através de programas específicos. As avaliações da produção científica têm sido usadas ainda para a construção de rankings de prestígio nacional e internacional para as instituições de ensino e pesquisa 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, no sentido utilizado por Bourdieu, conferindo prestígio aos pesquisadores, programas e instituições ranqueadas (Bourdieu, 2004).

Neste ensaio tratamos de diferentes aspectos relacionados ao uso dos artigos científicos como unidade de análise da produção científica e dos próprios periódicos científicos, utilizados para a divulgação do conhecimento produzido, como indicadores indiretos da qualidade dessa produção. Apesar da produção científica se traduzir em diferentes produtos como livros, comunicações em congressos, produtos técnicos de diferentes naturezas, focamos nos artigos científicos que tem sido universalmente utilizados, tanto para análises quantitativas quanto qualitativas da produção.

O ensaio foi construído a partir da análise da literatura pertinente e da experiência da autora em processos avaliativos da política científica, avaliação de programas de pós-graduação *estricto senso* e na atividade editorial em periódicos científicos publicados no país e no exterior.

2 O ARTIGO E A AVALIAÇÃO DA PRODUÇÃO CIENTÍFICA

A publicação de artigos científicos tem primordialmente dois objetivos principais para seus autores: a comunicação de seus resultados à comunidade de pares, essencial aos processos de construção de verdade e fonte do reconhecimento do mérito acadêmico, capital simbólico fundamental no

campo científico; e, garantir a autoria e a precedência nas descobertas (Barata, 2019). Entretanto, na atualidade, as exigências dos diferentes sistemas avaliativos aos quais o pesquisador se vê submetido, conferem aos artigos publicados papel central na tomada de decisão sobre financiamento à pesquisa, concessão de bolsas, progressão na carreira acadêmica, apoio a eventos, prêmios e outras formas de reconhecimento como os adicionais salariais por desempenho, adotados em algumas instituições acadêmicas.

A maior parte, senão a totalidade dos sistemas acadêmicos de avaliação têm nos artigos científicos ou tecnológicos, elemento chave nos seus processos avaliativos colocando, então o problema de como avaliá-los. Apenas o volume de publicações é insuficiente para orientar as avaliações e consequentes decisões das agências. Há duas alternativas: a classificação da produção segundo características formais dos artigos ou a avaliação qualitativa de uma amostra dos artigos produzidos, seja através de peer review pós-publicação, seja através da contagem de citações recebidas.

Muitas classificações foram e são adotadas no sentido de categorizar a produção em artigos científicos: publicações em inglês ou na língua nativa, cooperação nacional ou internacional entre os autores, ordem de autoria, resultados de projetos com ou sem financiamento de agências de fomento, publicação com ou sem alunos de iniciação científica, mestrado ou doutorado, publicações vinculadas ou não às linhas de pesquisa do pesquisador ou do programa no qual ele atua. Enfim, são quase infinitas as possibilidades de classificação a partir de características formais dos artigos.

A avaliação por pares, ou o *peer review* de artigos é uma atividade frequente já que faz parte do processo de publicação científica e os cientistas tanto se submetem a ela quando se trata de publicar os próprios trabalhos, quanto executam a mesma tarefa como parte de seu compromisso com os outros membros da comunidade, possibilitando a seleção dos trabalhos para publicação em diferentes veículos. Geralmente essa atividade é realizada anteriormente à aceitação e como parte do processo de julgamento da qualidade dos artigos a serem publicados (Wainer; Eckman; Rocha, 2015).

Porém, com o enorme crescimento da produção surgem novas modalidades de revisão por pares seja antes da publicação, nos repositórios de *preprints*, seja posteriormente à publicação quando editores selecionam os

artigos que eles consideram interessantes em cada número de um periódico (escolhas do editor) ou ainda quando algumas revistas escolhem o melhor do ano entre os artigos publicados. Há ainda iniciativas como a *Faculty of 1000* (F1000) que seleciona e recomenda os artigos mais notáveis na literatura biomédica a partir da opinião de um conjunto de especialistas (Wainer; Eckman; Rocha, 2015).

A revisão por pares, entretanto, tem sido objeto de muito criticismo quanto ao grau de objetividade e a endogenia (pouca diversidade de gênero, nacionalidade, instituições etc.) que predomina nos comitês de especialistas. Oppenheim (1997) demonstrou que os resultados do *Research Assessment Exercise* de 1992 e 1996 realizado para identificar o desempenho de departamentos acadêmicos britânicos, com base em informações de financiamento da pesquisa, número de alunos titulados, número de pesquisadores ativos e inativos e publicações entre 1988 e 1992, apresentou correlação forte com a contagem do número de citações recebidas pelos pesquisadores no mesmo período. A posição obtida pelos departamentos de uma mesma área de conhecimento foi correlacionada com a posição produzida pelo número total de citações dos professores e pelo número médio de citações por docente. Todas as correlações foram significantes. O autor chama a atenção para o fato de que os achados não significam que alto número de citações corresponda a alta qualidade, simplesmente há correlação entre os *rankings* computados a partir de citações com aqueles produzidos através da avaliação por pares.

Szkló e Nieto (2019) discutem o uso bastante frequente do coeficiente de correlação enquanto indicativo de correlação linear entre dois conjuntos de medidas lembrando que não necessariamente a associação linear é uma boa medida da concordância entre duas classificações. Pode haver correlação linear perfeita mesmo que exista viés entre as medidas correspondentes nas duas séries, ou seja, se sistematicamente uma das medidas é mais alta ou mais baixa que a outra a concordância não será tão boa mesmo que a correlação seja perfeita. Essa limitação se aplica tanto para a comparação entre valores em escala contínua (Pearson) quanto para escalas ordinais (Spearman). Portanto, o fato de haver correlação entre número de citações e avaliação por pares não significa que as duas

classificações estejam medindo os mesmos aspectos do desempenho acadêmico.

As medidas baseadas em citações são vistas como indicativas da visibilidade e do impacto da pesquisa enquanto a avaliação de qualidade implica um conceito mais amplo que envolve diferentes aspectos exigindo um *mix* de diversos critérios. Diversos autores veem os indicadores bibliométricos como complementares à revisão por pares e reconhecem que necessariamente haverá algum grau de correlação entre ambos, mas não esperam que haja “concordância ortogonal” entre eles, ou seja, sempre haverá um grau de autonomia entre as diferentes abordagens (Barreto, 2013a; Barreto, 2013b; Rinia *et al.*, 1998; Struchiner, 2013).

Waltman e Costas (2013) compararam as avaliações dos revisores no F1000 com as citações recebidas pelos artigos para verificar em que medida a qualidade avaliada pelo *peer review* é capaz de prever as citações que um artigo receberá. O F1000 é um serviço comercial online de revisão por pares pós publicação para pesquisas nas áreas biológicas, biomédicas e médicas. Ele conta com mais de 5.000 revisores que selecionam artigos interessantes lidos por eles. Os artigos são classificados em bons, muito bons ou excepcionais. No início de 2013 já havia 132.622 recomendações no site cujos artigos foram classificados como bons (58%), muito bons (35%) ou excepcionais (7%). Entre 2002 e 2013 esses percentuais se mantiveram relativamente constantes. Cerca de 80% dos artigos foram recomendados até o quarto mês após sua publicação embora o repositório também conte com artigos antigos considerados relevantes pelos revisores.

Waltman e Costas (2013) identificaram 38.327 artigos publicados entre 2006 e 2009 e indexados na *Web of Science* (WoS). Estes artigos foram comparados aos 1,7 milhões de artigos dessa base no mesmo período, para a análise do número de citações. Para os artigos que não receberam recomendações o número médio de citações foi 7,2 considerando uma janela de citações de 3 anos após a publicação. Os artigos classificados como bons apresentaram em média 20,7 citações, enquanto os muito bons obtiveram 37,6 citações em média e os excepcionais, 68,6. Assim, embora a média não seja uma boa medida para o número de citações, os dados sugerem que as recomendações e o número de citações recebidas são coerentes entre si. No

entanto, os autores também observaram que muitos dos artigos altamente citados não haviam recebido nenhuma recomendação. Os resultados mostram que existe coerência entre as recomendações dos *experts* e o número de citações recebidas sugerindo que as duas classificações não são independentes, porém, o fato de artigos muito citados não terem recebido indicações sugere também algum grau de autonomia entre as duas abordagens.

Bornmann e Leydesdorff (2013) retomam a discussão lembrando que qualidade é um atributo complexo que envolve além do impacto outros fatores como importância, originalidade, rigor e coerência interna. Os autores calcularam a correlação de Spearman entre diversos indicadores de citações normalizados pelo campo de conhecimento e a avaliação de artigos pelos especialistas que compõem o F1000. Eles observaram, para a análise de 125 artigos publicados entre 1981 e 2010 que os indicadores bibliométricos que melhor se correlacionaram com os scores de avaliação foram o percentil de citações na área de conhecimento e a razão entre citações observadas e esperadas. Mesmo assim, esses indicadores explicam apenas 20% da variação observada quanto à classificação de qualidade.

Além da qualidade do artigo baseada no mérito acadêmico, ou seja, na correção dos aspectos metodológicos, as recomendações dos *experts* certamente levaram em conta aspectos relativos à relevância científica dos mesmos e eventualmente à relevância social e potencial impacto dos resultados para a solução de problemas, questões essas, cada vez mais centrais para a política científica (Scott, 2007).

Desde a última década do século XX, a política científica tem sido objeto de maior atenção por diversos setores da sociedade enfatizando diferentes aspectos da relevância da ciência. As pressões políticas mais constantes sobre o campo científico apontam a necessidade dos produtos científicos mostrarem aplicação, além de exigirem a prestação de contas pelo investimento público feito em ciência. Outros aspectos bastante enfatizados dizem respeito à maior democratização na definição de prioridades e à cobrança de resolução de problemas sociais cada vez mais complexos (Scott, 2007).

Essa tensão entre as demandas da sociedade e a necessidade de preservação da autonomia por parte da comunidade científica, já apontado por Bourdieu (2004) na análise do campo científico, torna-se cada vez mais pronunciada assumindo graus diferenciados de questionamento e até de descrédito progressivo da contribuição da ciência para o desenvolvimento social. Segundo aponta Demeritt (2000) a posição anterior, em que a ciência era vista como um bem público como um fim em si mesma, rapidamente vai sendo substituída por uma visão utilitária do significado da ciência e da técnica para os processos de produção pós-industriais, tão importantes para as corporações quanto para os governos. Nesse processo a autonomia acadêmica vai progressivamente perdendo graus de liberdade e a disputa por recursos públicos se torna cada vez mais dependente da demonstração da relevância econômica e social dos resultados da ciência.

Há uma constante contradição no interior da comunidade científica entre a recusa à metáfora da torre de marfim e a relativa perda de autonomia subjacente às demandas crescentes da sociedade e do Estado, via agências de fomento, por maior impacto e inovação. Nesse ambiente em permanente transformação, os cientistas sociais, mas principalmente os cientistas do campo das ciências humanas, são os que vocalizam de maneira mais crítica a oposição entre uma prática reflexiva e crítica que segundo eles estaria na essência da docência e da produção acadêmica e um compromisso mais utilitário e instrumental predominante nas relações entre universidade, Estado e sociedade (Chauí, 2017; Demeritt, 2000; Scott, 2007).

De todo modo, nenhum dos pesquisadores propõem que os sistemas avaliativos deixem de considerar a relevância científica ou social da ciência. A relevância científica é um item menos discutido embora fundamental. Ela diz respeito aos motivos de ordem teórica e prática que tornam a pesquisa importante, isto é, quais são as contribuições teóricas e/ou metodológicas que a pesquisa traz para o campo do conhecimento no qual foi realizada. Muitas vezes ela se confunde com o mérito ou rigor acadêmico mas não se trata do mesmo conjunto de aspectos, uma vez que, pesquisas absolutamente corretas do ponto de vista do seu desenvolvimento podem ser irrelevantes, seja porque as perguntas de investigação não contém nenhum aspecto de originalidade ou inovação, seja porque os métodos empregados são convencionais não

conduzindo ao questionamento de verdades já estabelecidas, seja porque os resultados não acrescentam nada de novo ao conhecimento já existente, constituindo mais uma pesquisa de confirmação do que propriamente de investigação.

Kuhn (1989) menciona que o trabalho do cientista no interior dos paradigmas da ciência normal devota-se, em grande parte, a aumentar a precisão e acurácia ou ainda a extensão dos conhecimentos já sancionados pelo paradigma. Há ainda a possibilidade de aplicar-se a fenômenos menos interessantes ainda não totalmente explorados ou buscar a resolução de ambiguidades residuais, ainda não devidamente compreendidas e explicadas pelo paradigma vigente. Assim, apenas uma pequena parte da produção científica conteria aspectos de originalidade e inovação.

Somada a essas características do trabalho científico, o fato de que parte da produção acadêmica decorre das atividades de formação de novos cientistas nos programas de pós-graduação *estricto senso*, para os quais os prazos de conclusão são previamente estabelecidos dentro de margens relativamente estreitas, não é de se estranhar que parte considerável do que é produzido traga pouca inovação e originalidade.

Stokes (1997) apresenta em seu livro "*Pasteur's Quadrant*" uma solução esclarecedora para a tensão entre as questões do conhecimento por um lado, e as questões práticas, por outro. Ao contrário da crença dominante durante a maior parte do século XX de que o processo de produção da ciência é predominantemente linear, evoluindo por acréscimos ao conhecimento, que acabam por resultar em aplicações e novas tecnologias, Stokes postula que existem basicamente três tipos de pesquisa, que desde a origem são distintas entre si, e que não há obrigatoriamente, nem majoritariamente, continuidade entre elas: a pesquisa básica dirigida à solução de problemas teóricos e conceituais do conhecimento dos fenômenos e que não tem por objetivo a aplicação prática desses conhecimentos; a pesquisa tecnológica que é claramente dirigida por problemas práticos visando aplicar o conhecimento científico em processos e artefatos de caráter utilitário; e, a pesquisa estratégica que desde o princípio está dirigida aos dois polos opostos: o acréscimo ao conhecimento sobre leis e fenômenos e a solução de problemas práticos.

Necessariamente a pesquisa estratégica coloca tanto a questão do rigor e da relevância científica quanto a questão da relevância social. A pesquisa em saúde é predominantemente desenvolvida nesse espaço que o autor designou de Quadrante de Pasteur ressaltando essa combinação de objetivos por oposição, de um lado ao Quadrante do Bohr, ocupado pela pesquisa básica e ao Quadrante de Edson, ocupado pela pesquisa tecnológica.

Dobrow e colaboradores (2017) afirmam que na pesquisa em saúde a preocupação com a relevância é lugar comum, mas que mesmo assim, não há uma concepção objetiva do que se entende por relevância. De certo modo, a questão da relevância se contrapõe à preocupação excessiva com o rigor e a validade interna das pesquisas muitas vezes às custas de redução da validade externa e perda do potencial de aplicação. Os autores chamam a atenção para o quanto a discussão sobre a relevância se associa a esta perspectiva de valorização da validade externa que, em certa medida se viu comprometida pela vaga da medicina baseada em evidências, na qual as evidências ficaram reduzidas àquelas oriundas de revisões sistemáticas seguidas de meta análise e aos resultados de ensaios controlados e randomizados, procedimentos que geralmente comprometem de maneira importante a validade externa dos resultados. Além da recuperação da validade externa, os autores apontam também o movimento da pesquisa translacional com seus modelos de tradução do conhecimento científico para as aplicações práticas tanto na clínica quanto na saúde pública, como decorrentes da preocupação atual com a relevância.

Guimarães (2011) associa o crescimento da pesquisa estratégica à necessidade de mediação entre universidade e outros setores sociais visto que não apenas a formulação das demandas para o desenvolvimento científico, mas também o delineamento de novas diretrizes para a formação de pesquisadores torna-se cada vez mais relevante. A abertura para o mundo extrauniversitário é importante para aproximar a pesquisa científica do cenário de aplicação, modificando assim, não apenas as questões de investigação, mas principalmente modelando os produtos da pesquisa para aplicações mais ou menos imediatas à solução de problemas relevantes do ponto de vista do desenvolvimento econômico e social.

Entretanto, a relevância da pesquisa não pode ser reduzida à pesquisa instrumental. Ela está relacionada à *phronesis*, ou seja, a capacidade de agir de maneira flexível e adaptativa frente a situações incertas, em contextos diversos e cheios de surpresa. Ela se baseia na experiência acumulada e no julgamento profissional, no conhecimento prático necessário para lidar com os desafios de contextos particulares. A pesquisa relevante, portanto, será aquela capaz de se valer desse conhecimento prático acumulado articulando essa experiência aos conhecimentos científicos capazes de solucionar os problemas a partir de um patamar de reflexividade e conhecimento lógico que ultrapassa o senso comum (Scaratti *et al.*, 2017).

Assim, a relevância pode ser definida como:

[...] alguma coisa que importa para a experiência cotidiana de trabalho dos profissionais: para seus problemas e questões, para os significados e valores que eles atribuem aos problemas que enfrentam e para o conhecimento que auxilia na análise, compreensão e desenvolvimento de possíveis soluções (Scaratti *et al.*, 2017, p. 59).

Já o impacto se refere a:

[...] transformações geradas pelos processos de produção de conhecimentos, tanto em relação à possibilidade de mudança quanto à expansão da aprendizagem, em cada campo de atuação. O conhecimento tem impacto quando ilumina contradições e dilemas, desafia as certezas e alimenta processos reflexivos capazes de suportar ações voltadas para a mudança do sistema de atividades (Scaratti *et al.*, 2017, p. 59).

A preocupação crescente com a relevância e o impacto social das atividades de pesquisa pode ser evidenciada em dois movimentos, relativamente recentes, um no campo da formação representado pela valorização dos programas de pós-graduação profissional, principalmente os doutorados; e, a chamada pesquisa de implementação, um subcampo da pesquisa avaliativa que está voltada para o acompanhamento e a avaliação da efetividade na implementação de conhecimentos científicos à prática (Geng; Peiris; Kruk, 2017).

Grande parte das agências de fomento introduziram em seus processos avaliativos itens relativos à relevância social dos projetos, no entanto, ainda não existem instrumentos em que o conceito possa ser objetivado (Reale *et al.*, 2018; Wilbertz, 2013). No caso do Brasil, tanto o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) quanto as Fundações de Amparo à Pesquisa dos estados, solicitam dos proponentes a explicitação dos

potenciais usos dos resultados da pesquisa e, aos avaliadores, que considerem esses aspectos na decisão. Do mesmo modo a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), inclui como um dos tópicos da avaliação dos programas de pós-graduação a inserção social, tanto em termos regionais como nacionais ou internacionais. Porém nenhuma dessas agências dispõem de orientações claras e objetivas relativas a como tais aspectos devem ser avaliados.

Quanto aos artigos científicos não parece haver ainda, por parte dos periódicos, orientação para que os revisores considerem, além do rigor, qualidade acadêmica, originalidade e contribuição para o avanço do conhecimento, os aspectos de aplicabilidade ou inovação dos resultados apresentados.

3 O PERIÓDICO CIENTÍFICO E A AVALIAÇÃO DA PRODUÇÃO CIENTÍFICA

Além de possíveis classificações baseadas em características dos artigos existem outras baseadas nas características dos periódicos nos quais os artigos foram publicados. O CNPq, por exemplo, nas suas avaliações dos grupos de pesquisa, classifica os artigos segundo a publicação em periódicos nacionais (publicados em português) ou estrangeiros (erroneamente designados de internacionais e publicados em outras línguas) sem explicitar o significado intrínseco à classificação, uma vez que existem periódicos nacionais com circulação internacional e periódicos estrangeiros com circulação exclusivamente local. Outros critérios que são empregados em concursos acadêmicos ou seleção de projetos, separam os artigos de acordo com a publicação em periódicos com ou sem revisão por pares, ou periódicos indexados ou não indexados.

As classificações baseadas nos periódicos e não em cada um dos artigos surgem como alternativa à simples quantificação do volume de produção, por um lado, e, à impossibilidade de avaliar qualitativamente um grande número de artigos. Essas classificações pressupõem que a aceitação de um artigo por um periódico indexado e com revisão por pares garante de certo modo sua qualidade (Barata, 2016).

Assim, à medida em que as avaliações da produção científica vão se valendo dos artigos científicos para a construção de indicadores de

desempenho, as bases de indexação vão se tornando importantes não apenas como meio de circulação dos conteúdos, mas também como fornecedoras de medidas bibliométricas relativas ao número de citações, acessos, downloads, compartilhamentos e mais recentemente das medidas alternativas como as altimetrias.

Minniti e colaboradores (2018) compararam as características dos artigos em acesso aberto, produzidos nos países latino-americanos, analisando a produção indexada no *Scientific Electronic Library Online* (SciELO) e na WoS, no período de 2005 a 2017. Dos 983 mil documentos procedentes da América Latina indexados na WoS, 212 mil (21,6%) estão em *open access* enquanto na SciELO esses documentos são mais numerosos atingindo 387 mil. Em ambas as bases a grande maioria dos artigos foram publicados pelos países do Cone Sul, principalmente pelo Brasil. Há no entanto maior representação relativa dos países andinos e do Caribe na base da SciELO.

Analisando a tendência temporal, houve crescimento acentuado de artigos em *open access* na WoS a partir de 2014, enquanto no mesmo período houve declínio na indexação na SciELO sugerindo que os editores, diante da possibilidade de indexar os periódicos em uma base internacional, preferem essa alternativa à indexação em uma base regional. Quanto à língua de publicação, para o total de documentos indexados na WoS, 85% estavam publicados em inglês, 7,4% em espanhol e 7,4% em português. Considerando apenas os artigos brasileiros esses percentuais sofrem grande alteração quando se consideram os artigos em acesso aberto, passando para 70% em inglês, 1,2% em espanhol e 29% em português na WoS e 38% em inglês, 2% em espanhol e 60% em português na SciELO

3.1 Classificação Qualis

O *Qualis Periódicos* foi instituído pela CAPES em 1998 como forma de classificar os periódicos científicos nos quais era divulgada a produção científica de docentes e discentes dos programas de pós-graduação *estricto sensu*. Diante da necessidade de considerar na avaliação dos programas, não apenas o número de artigos produzidos, mas também a qualidade desses artigos e, na impossibilidade de avaliar cada artigo em si, o Qualis foi pensado como alternativa viável a partir da ideia de classificar os veículos confiando no

processo de *peer review* por eles realizado (Barata, 2015; Barata, 2016; Ferreira; Moreira, 2002).

Desta maneira, a qualidade dos artigos poderia ser aferida indiretamente baseando-se nas características dos periódicos nos quais os artigos haviam sido publicados. A premissa subjacente era que os periódicos com maior circulação e impacto possuíam processos de revisão por pares caracterizados por avaliação rigorosa do mérito acadêmico, consideração da relevância científica e competição, resultando em seleção dos melhores artigos.

Este modelo de publicação atendia às características da divulgação científica dominantes nas áreas de Ciências Exatas e da Terra, Engenharias, Ciências Agrárias, Ciências Biológicas e Ciências da Vida, porém não correspondia plenamente ao que era praticado nas Ciências Humanas e Sociais Aplicadas ou nas áreas de Artes, Literatura e Linguística.

Desse modo, a operacionalização do Qualis Periódicos, tanto em sua versão inicial quanto nas subsequentes, foi definida em cada área de avaliação com base em diferentes critérios classificatórios que iam desde o uso exclusivo de determinadas medidas bibliométricas como o fator de impacto, até a combinação de características formais das publicações e avaliação da importância do periódico para o campo científico segundo apreciação dos comitês de avaliadores (Barata, 2016).

Facilmente se deduz que, desse modo o Qualis Periódico poderia ser usado para avaliar o desempenho comparativo entre programas de pós-graduação pertencentes a uma mesma área de avaliação, mas não serviria como instrumento para mensurar a produção científica em si, visto que, os critérios usados pelas áreas não apresentavam nenhuma equivalência e as travas impostas à classificação nos diversos estratos dificultavam uma apreciação adequada da produção científica fora do contexto de avaliação dos programas.

Na versão atual adotada pela CAPES não são mais aplicadas travas e ao invés de sete estratos são utilizados oito estratos (quatro estratos A e quatro estratos B), cada um correspondendo a 12,5% da distribuição dos indicadores bibliométricos normalizados em cada área do conhecimento. Além disso cada área classifica apenas um conjunto de periódicos baseado no número de

artigos publicados pela área de avaliação no período considerado. Esse é o ponto fraco da nova versão, uma vez que gera enorme instabilidade à classificação pois ao invés de distribuir os periódicos pelas áreas de conhecimento, adota um critério operacional destituído de valor intrínseco, baseado exclusivamente no número de artigos publicados pelas diferentes áreas de avaliação, no período avaliativo, número que poderá se alterar no período seguinte. Por exemplo, a revista *Lancet*, talvez o mais tradicional e antigo periódico científico da área médica, passa a ser considerado na lista da Saúde Coletiva simplesmente porque essa área de avaliação detém o maior número de artigos nela publicados nos últimos quatro anos.

Para que o Qualis Periódicos pudesse servir como ferramenta de avaliação da produção científica brasileira, vinculada ou não aos programas de pós-graduação, seria necessário introduzir algumas mudanças na sua construção atendendo às seguintes características:

- a) A classificação deveria ser aplicada exclusivamente a periódicos científicos diferenciando-se claramente estes das publicações técnicas e dos periódicos de popularização da ciência.
- b) A classificação deveria ter um referencial mundial e não apenas estar baseada, retrospectivamente, naquilo que foi produzido pelos programas de pós-graduação. Os pontos de corte dos indicadores para definição dos estratos deveriam utilizar toda a coleção, por exemplo da base *Scopus*, em cada área do conhecimento considerada.
- c) As classificações nas várias áreas de conhecimento deveriam se equivaler, ou seja, a classificação deveria ter algum grau de comensurabilidade entre as áreas.

A classificação do Qualis deve estar apoiada em dois princípios fundamentais: a circulação do conhecimento produzido e o impacto da publicação na geração de novos conhecimentos. A circulação de um periódico e do conteúdo por ele veiculado depende do formato no qual o mesmo está à disposição na *web* e também dos mecanismos de busca disponíveis para acessá-lo.

Com o crescimento exponencial dos periódicos, o papel das bases indexadoras, se mantém em alta na comunidade científica visto que a seleção de periódicos para integrá-las, ainda que influenciado por diversos fatores nem sempre diretamente relacionados com a relevância e o mérito acadêmico, acaba por funcionar como um filtro de qualidade. Dependendo da cobertura e abrangência dessas bases, em alguns campos científicos praticamente não há consulta a periódicos de não estejam ali indexados. Ou seja, o conhecimento que importa deve ser encontrado em periódicos contidos por exemplo na WoS ou na base Scopus. Para aqueles campos de conhecimento básico tanto das ciências exatas quanto das ciências biológicas, mesmo uma base mais restrita pode ser suficiente.

Para áreas de conhecimento mais aplicadas como são as engenharias, as ciências agrárias e as ciências da saúde, bem como algumas das ciências sociais, quanto mais ampla a base de indexação maior a probabilidade de acessar conhecimento relevante. Finalmente para diversas ciências humanas e para a área de artes e linguística/ literatura, um buscador poderoso, com ampla base de consulta, como o *Google Scholar* representa o recurso mais valioso para a busca e localização de livros e periódicos (Barreto, 2013a). Os repositórios também são mecanismos importantes de acesso e circulação da produção científica dependendo, no entanto de contarem com mecanismos de busca realmente eficientes, além do preparo adequado dos documentos para garantir sua recuperação.

Periódicos que não podem ser encontrados através da Internet podem até ser de qualidade, assim como os artigos nele veiculados, no entanto tem probabilidade muito pequena ou nula de serem conhecidos e utilizados por parcela significativa dos pesquisadores. Desse modo não cumprem adequadamente com a exigência de auxiliar na circulação do conhecimento e, portanto, não devem ser valorizados em uma ferramenta destinada a avaliar a produção científica.

Para que cada periódico tenha uma única classificação é necessário distribuir os periódicos de acordo com a área de conhecimento ao qual ele se dedica. Geralmente as bases de indexação e catálogos utilizam diferentes classificações, exigindo então um trabalho prévio de normalização e

correspondência à árvore de conhecimento usada pelas agências de fomento brasileiras.

Embora a listagem dos periódicos possa continuar sendo elaborada a partir dos registros de produção dos programas, a estratificação deveria adotar um referencial mundial. Do ponto de vista operacional, a base *Scopus*, permite obter os indicadores normalizados segundo cada uma das áreas e subáreas de indexação dos periódicos, para todas as áreas representadas na base, fundamentalmente as ciências exatas, biológicas, engenharias, ciências agrárias e ciências da saúde, e algumas das ciências sociais aplicadas e ciências humanas.

O último requisito para que o Qualis possa se constituir em ferramenta de avaliação da produção científica em geral é estabelecer uma classificação que seja semelhante entre as áreas de conhecimento, ou seja, que os periódicos classificados no estrato A1 possam se equivaler em todas as áreas quanto à circulação e probabilidade de citação.

Como todas as medidas bibliométricas sofrem a influência de inúmeros fatores como língua, tamanho da comunidade acadêmicas, modos de produção do conhecimento científico em cada áreas, dentre outros, não é possível estabelecer pontos de corte da classificação em termos absolutos (Glänzel; Moed, 2002).

Para garantir comensurabilidade entre as classificações os limites dos estratos não podem estar baseados em valores absolutos dos indicadores bibliométricos. O ideal seria adotar processos de normalização dos valores ou a distribuição em percentis, que é exatamente o que está previsto no Novo Qualis (Bornmann, 2013).

Com essas modificações o Qualis Periódico pode tanto servir ao seu propósito inicial de ferramenta para avaliar o desempenho dos programas de pós-graduação quanto como instrumento para a avaliação da produção científica em geral. Um Qualis Periódicos assim construído tem maior clareza de critérios, é mais preciso, uma vez que cada área tem um menor número de títulos para analisar e estratificar; o uso de medidas bibliométricas torna a classificação mais objetiva e exata e, a adoção de percentis pré-estabelecidos permite a comparação direta entre as diferentes áreas de conhecimento. As mudanças adotadas pela CAPES atendem à maioria desses

princípios fazendo crer que a classificação estará mais próxima de representar um bom instrumento para avaliação da produção.

3.2 Medidas de impacto bibliométrico dos periódicos

O impacto bibliométrico dos periódicos pode ser mensurado através de diversas medidas que funcionam como *proxy* da visibilidade e da utilização dos artigos para a elaboração de novas pesquisas. Medidas bibliométricas como o Fator de Impacto podem ser vistas como indicativas da probabilidade de um artigo vir a ser citado. Estudo realizado com artigos duplicados publicados em periódicos com diferentes fatores de impacto demonstrou que o fator de impacto da revista influi positivamente no número de citações recebidas pelo artigo. Artigos publicados em periódicos com fatores de impacto mais alto (acima de 1,11 em média) recebem o dobro de citações das suas cópias publicadas em periódicos com fatores de impacto menores do que 0,5 (Larivière; Gingras, 2010).

Cada uma das medidas de impacto bibliométrico possui limitações amplamente discutidas na literatura e assim, o uso de mais de uma medida pode ser interessante para compensar as deficiências entre elas. Dentre as medidas disponíveis as mais facilmente utilizáveis e compreensíveis são o fator de impacto ou o *cites per doc* (que é a medida equivalente na base *Scopus*), o índice *h*, o *SCImago Journal Rank* (SJR) e o índice *h5* do *Google Scholar*. Qualquer uma delas, ou um conjunto delas, pode ser usada para classificar os periódicos.

Há uma boa consistência entre essas diferentes medidas bibliométricas embora elas correspondam a três tipos de medidas distintas. O fator de impacto e o *cites per doc* se equivalem variando apenas em função da amplitude de cobertura das bases e da definição de documento citável.

O índice *h* e o índice *h5* também se equivalem variando apenas em função do tempo e da amplitude de cobertura das bases. Ambos diferem do fator de impacto e do *cites per doc* pois ao invés de considerar a razão entre o número de citações recebidas durante dois anos pelos artigos publicados em determinado ano e o número de artigos publicados, o índice *h* indica o número de artigos que receberam *h* citações cada. O índice *h5* considera apenas os artigos publicados nos últimos cinco anos. Embora medindo

aspectos diversos há boa correlação entre o índice h e o fator de impacto ou cites per doc dos periódicos.

O SJR é a medida mais complexa e obtida pela ponderação das citações pelas características dos jornais citantes. Ela mede aspectos diferentes e assim apresenta correlação menor com as medidas anteriores. Portanto é possível construir um processo de classificação dos periódicos usando medidas objetivas estabelecendo combinações entre elas ou utilizando-as de modo complementar.

Alguns estudiosos da ciência da informação começam a discutir os impactos que as novas formas de divulgação do conhecimento científico possam ter sobre a cientometria. Tendo em vista que atualmente a informação científica é majoritariamente disseminada por meios eletrônicos há cada vez menor probabilidade dos autores lerem um periódico em sua integralidade. Habitualmente as revisões bibliográficas são conduzidas eletronicamente buscando tópicos particulares independentemente dos periódicos nos quais os artigos estejam publicados. Assim, os artigos podem ser recuperados, lidos e citados com base exclusivamente em seus méritos, seja qual for a reputação e o impacto do periódico que os veicula (Lozano; Larivière; Gingras, 2012).

Os argumentos apresentados são bons, entretanto ignoram que diante de um volume muito grande de referências recuperadas nas buscas eletrônicas, um recurso que os pesquisadores têm para selecionar o material mais relevante é justamente o prestígio dos periódicos nos quais os artigos estão publicados. Assim, acreditamos que uma ferramenta de avaliação baseada na classificação dos periódicos ainda tem utilidade, sempre que o volume do material a ser avaliado for muito grande para permitir a análise individualizada de artigo por artigo.

Resta um último ponto a considerar que diz respeito aos vieses de publicação já apontados quando diferentes bases indexadoras são analisadas. É preciso considerar que para cobrir de modo adequado a produção de uma dada área é necessário examinar não apenas a produção indexada nas bases de circulação internacional, mas incluir também bases regionais, normalmente mais dedicadas a ampliar a circulação da produção nas regiões que compartilham condições mais assemelhadas tanto na

problemática a ser estudada quanto em relação à política científica e tecnológica. Para a chamada ciência básica, talvez essa preocupação não seja relevante, mas sem dúvida deverá ser considerada nas avaliações envolvendo ciências aplicadas como é o caso de inúmeras áreas do conhecimento.

4 CONSIDERAÇÕES FINAIS

A avaliação da produção científica através dos artigos científicos publicados tem duas grandes vertentes principais: a avaliação dos artigos em si e a avaliação dos periódicos nos quais eles são publicados. Na avaliação dos artigos em si, podem ser usadas fundamentalmente três estratégias distintas e complementares: a classificação segundo características formais do artigo, a contagem das citações recebidas, e a avaliação qualitativa baseada nos procedimentos de peer review pós-publicação.

O procedimento mais simples, e mais automatizável é a contagem do número de citações, porém sabemos que as citações não refletem exclusivamente a qualidade, sendo medidas que se referem principalmente à recepção do artigo pela comunidade.

As classificações segundo características formais podem ser extraídas eletronicamente de bases de dados, desde que os artigos estejam indexados, mas embora sirvam para caracterizar certos padrões de publicação existentes nas áreas de conhecimento não indicam precisamente a qualidade dessa produção.

O *peer review*, além dos problemas inerentes a essa forma de avaliação tem a limitação de não poder ser aplicado à totalidade da produção, ficando restrito à análise de amostras aleatórias ou artigos selecionados pelos autores ou diretores das instituições avaliadas, dependendo do contexto de avaliação.

Na avaliação dos periódicos como indicadores indiretos da qualidade da produção, utilizam-se diferentes indicadores bibliométricos, seja como fator de ponderação dos artigos, seja como critérios para construir classificações como o Qualis periódicos da CAPES. Erroneamente a comunidade costuma confundir o Qualis como uma ferramenta de simples quantificação da produção, coisa que ele não é. O Qualis é um recurso indireto de avaliação

qualitativa da produção ao atribuir aos periódicos posições decorrentes do âmbito de circulação e do impacto (utilização dos artigos na construção de novos conhecimentos) e transferir essa qualificação a cada um dos artigos neles publicados.

Como em toda questão complexa, não pode haver uma única forma de realizar a avaliação e, o uso de diferentes estratégias pode contrabalançar limitações existentes em cada uma delas, dando uma perspectiva mais abrangente e multifacetada para os avaliadores.

REFERÊNCIAS

BARATA, R. B. A ABRASCO e a pós-graduação stricto sensu em saúde coletiva. In: LIMA N. T.; SANTANA J. P.; PAIVA C. H. A. (org.) **Saúde Coletiva: a ABRASCO em 35 anos de história**. Rio de Janeiro: Editora Fiocruz: ABRASCO, 2015. cap. 8, p.169-198.

BARATA, R. B. Desafio da editoração de revistas científicas brasileiras da área da saúde. **Ciência & Saúde Coletiva**, Rio de Janeiro, v. 24 n. 3, p. 929-939, 2019. DOI [10.1590/1413-81232018243.29952016](https://doi.org/10.1590/1413-81232018243.29952016).

BARATA, R. B. Dez coisas que você deveria saber sobre o Qualis. **Revista Brasileira de Pós-Graduação**, Brasília, v.13, n. 30, p.13-40, 2016. Disponível em: <https://rbpg.capes.gov.br/rbpg/article/view/947>. Acesso em: 31 out. 2023.

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. DOI <https://doi.org/10.1590/S0034-8910.2013047005073>.

BARRETO, M. L. Como avaliar as ciências com uma deficiente ciência da avaliação científica? (debate). **Cadernos de Saúde Pública**, Rio de Janeiro, v. 29, n. 9, p. 1719-1721, 2013b. Disponível em: <https://scielosp.org/pdf/csp/2013.v29n9/1719-1721/pt>. Acesso em: 31 out. 2023.

BORNMANN, L. How to analyze percentile citation impact data meaningfully in bibliometrics: the statistical analysis of distributions, percentile rank classes and top-cited papers. **Journal for the American Society for Information Science and Technology**, New York, v. 64, n. 3, p. 587-595, 2013. DOI <https://doi.org/10.1002/asi.22792>.

BORNMANN, L.; LEYDESDORFF, L. The validation of (advanced) bibliometric indicators through peer assessment: a comparative study using data from InCites and F1000. **Journal of Infometrics**, Taipei, v. 7, n. 2, p.286-291, 2013. DOI <http://dx.doi.org/10.1016/j.joi.2012.12.003>.

BOURDIEU, P. **Os usos sociais da ciência: por uma sociologia clínica do campo científico**. São Paulo: Editora UNESP, 2004. 85 p.

CHAUÍ, M. Avaliação irracional da pesquisa e eclipse da docência na USP. **Revista ADUSP**, São Paulo, v. 60, [s.n.], p. 54 – 65. 2017. Disponível em: <https://www.adusp.org.br/files/revistas/60/mat04.pdf>. Acesso em: 31 out. 2023.

DEMERRIT, D. The new social contract for science: accountability, relevance, and value in US and UK science and research policy. **Antipode**, Malden, v. 32, n. 3, p. 308-329. 2000. DOI <https://doi.org/10.1111/1467-8330.00137>.

DOBROW, M. J. *et al.* Understanding relevance of health research: considerations in the context of research impact assessment. **Health Research Policy and Systems**, Geneva, v. 15, [s.n.], p. 31-40, 2017. DOI <https://doi.org/10.1186%2Fs12961-017-0188-6>.

FERREIRA, M. M.; MOREIRA, R. L. **Capes 50 anos**: depoimentos ao CPDOC/FGV. Brasília: MEC, 2002.

GENG, E. H.; PEIRIS, D.; KRUK, M. E. Implementation science: relevance in the real world without sacrificing rigor. **PLoS Medicine**, San Francisco, v. 14, n. 4, 2017. DOI <https://doi.org/10.1371/journal.pmed.1002288>.

GLÄNZELL, W.; MOED, H. F. Journal impact measures in bibliometric research. **Scientometrics**, Budapest, v. 53, n. 2, p.171-193, 2002. Disponível em: <https://link.springer.com/article/10.1023/A:1014848323806>. Acesso em: 31 out. 2023.

GUIMARÃES, R. Desafios da pós-graduação em saúde humana no Brasil. **Revista de Saúde Pública**, São Paulo, v. 45, n. 1, p. 1-13, 2011. Disponível em: <https://www.scielo.br/j/rsp/a/p3GxTJhsdpXKtX3kdCc85k/?format=pdf&lang=pt>. Acesso em: 31 out. 2023.

KUHN, T. **A estrutura das revoluções científicas**. São Paulo: Editora Perspectiva, 1989. 218 p.

LARIVIÈRE, V.; GINGRAS, Y. The impact factor's Matthew effect: a natural experiment in bibliometrics. **Journal of the American Society for Information Science and Technology**, New York, v. 61, n. 2, p. 424-427, 2010. DOI <https://doi.org/10.1002/asi.21232>.

LOZANO, G. A.; LARIVIÈRE, V.; GINGRAS, Y. The weakening relationship between the impact factor and paper's citations in the digital age. **Journal of the American Society for information Science and Technology**, New York, v. 63, n. 11, p. 2140-2145, 2012. DOI <https://doi.org/10.1002/asi.22731>.

MINNITI, S.; SANTORO, V.; BELLI, S. Mapping the development of Open Access in Latin America and Caribbean countries: an analysis of Web of Science Core Collection and SciELO Citation Index (2005-2017). **Scientometrics**, Budapest, v. 117, [s.n.], p. 1905-1930, 2018. DOI <https://doi.org/10.1007/s11192-018-2950-0>.



OPPENHEIM, C. The correlation between citation counts and the 1992 research assessment exercise ratings for British research in genetics, anatomy and archaeology. **Journal of Documentation**, v. 53, n. 5, p. 477-487, 1997. DOI <https://doi.org/10.1108/EUM0000000007207>.

REALE, E. *et al.* Topics of social relevance in research project funding instruments: evidences from European countries. In: INTERNATIONAL CONFERENCE ON SCIENCE AND TECHNOLOGY INDICATORS, 23., 2018, Leiden. **Proceedings...** Leiden: [s.n.], 2018. Disponível em: <https://scholarlypublications.universiteitleiden.nl/handle/1887/65197>. Acesso em: 31 out. 2023.

RINIA, E. J. *et al.* Comparative analysis of a set of bibliometric indicators and central peer review criteria. Evaluation of condensed matter physics in the Netherlands. **Research Policy**, [s.l.], v. 27, [s.n.], p.95-107, 1998. DOI [https://doi.org/10.1016/S0048-7333\(98\)00026-2](https://doi.org/10.1016/S0048-7333(98)00026-2).

SCARATTI, G. *et al.* The social relevance and social impact of knowledge and knowing. **Management Learning**, [s.l.], v. 48, n. 1, p. 57-64, 2017. DOI [10.1177/1350507616680563](https://doi.org/10.1177/1350507616680563).

SCOTT, A. Peer review and the relevance of science. **Futures**, [s.l.], v. 39, n. 7, p. 827-845, 2007. DOI <https://doi.org/10.1016/j.futures.2006.12.009>.

STOKES, D. E. **Pasteur's Quadrant: basic science and technological innovation**. Washington: Brookings Institution Press, 1997. 248 p.

STRUCHINER, C. J. Avaliação da qualidade da produção científica e suas consequências imprevistas e indesejadas: um conceito auto evidente? (debate) **Cadernos de Saúde Pública**, Rio de Janeiro, v. 29, n. 9, p. 1716-1717, 2013. DOI <http://dx.doi.org/10.1590/0102311XCO040913>.

SZKLO, M., NIETO, F.J. **Epidemiology. Beyond the basics**. Burlington: Jones & Bartlett Learning, 2019.

WAINER, J.; ECKMAN, M.; ROCHA, A. Peer-selected "best papers": are they really that good? **PLoS One**, San Francisco, v.10, n. 3, 2015. DOI <https://doi.org/10.1371/journal.pone.0118446>.

WALTMAN, L.; COSTAS, R. F1000 recommendations as a potential new data source for research evaluation: a comparison with citations. **Journal of the American Society for information science and technology**, [s.l.], v. 60, n. 11, p. 2169-2385, 2013. DOI <https://doi.org/10.1002/asi.23040>.

WILBERTZ, J. **Evaluating societal relevance of research**. [s.l.]: University of Groningen, 2013. 29 p. Disponível em: <https://research.rug.nl/>. Acesso em: 03 jul. 2019.