

A necessária e difícil reforma da avaliação da investigação

Eloy Rodrigues ORCID: 0000-0002-7862-2681

'Not everything that can be counted counts and not everything that counts can be counted'
William Bruce Cameron (Informal Sociology: A Casual Introduction to Sociological Thinking. Random House, 1963).

'As the growth of knowledge is not autonomous, not random nor guided by the legendary 'invisible hand', it must be possible to improve the impact of science by cognizant governance to aim at better alignment of the research enterprise with our major national and global societal needs'

Frank Miedema (Open Science: the Very Idea. Springer, 2022)

1 - Introdução

A avaliação é uma componente essencial e determinante do sistema científico e académico. Como acontece em várias outras áreas da sociedade e da economia, a avaliação das atividades, dos resultados obtidos, e das pessoas e instituições que realizam investigação, não apenas valoriza o desempenho passado, como, ao fazê-lo (premiando ou “punindo”), determina em grande medida o futuro.

A avaliação da investigação dirige e condiciona a alocação de financiamento às instituições, projetos e investigadores, e influencia, direta ou indiretamente, a definição das áreas e tópicos de investigação prioritários, ou seja, a definição da agenda de investigação. Não menos importante é o impacto da avaliação na carreira académica e científica das pessoas que realizam investigação, nomeadamente nas que trabalham no contexto universitário. Por ser reconhecidamente determinante para o futuro individual dos professores e investigadores universitários (condicionando a sua progressão, o reconhecimento pelos pares e, mais prosaicamente, os seus rendimentos no final do mês), a avaliação molda os seus comportamentos, que naturalmente se ajustam aos parâmetros valorizados pelo sistema.

Não obstante existirem inúmeras variantes e especificidades (nacionais, institucionais e disciplinares) nos sistemas de avaliação utilizados, nas últimas décadas registou-se uma forte convergência para basear a avaliação científica e académica na utilização de métricas associadas à produção científica (número de publicações) e ao seu impacto (número de

citações). Nesse sentido, não parece incorreto falar de um sistema de avaliação da investigação global, que teve origem no mundo anglo-saxónico a partir dos anos de 1970, e se estendeu a todo o mundo desde o final do século passado.

A utilização generalizada, e em alguns casos exclusiva, dos indicadores bibliométricos na avaliação e, por outro lado, a afirmação dos conceitos e das práticas do acesso aberto e da ciência aberta, convergiram na última década para a formação de uma visão crítica sobre o sistema de avaliação dominante. A consciência dos problemas e limitações da atual avaliação da investigação, e da necessidade de a reformar profundamente, tem vindo a ganhar maior expressão, quer na comunidade científica e nas suas instituições, quer ao nível político. No entanto, o avanço da reforma tem sido ainda muito limitado, ao nível da profundidade sistémica das mudanças, e do seu alcance, ou seja, do número de instituições académicas e de agências de financiamento, que já alteraram significativamente as suas práticas.

Neste texto iremos descrever sumariamente as origens do atual sistema de avaliação da investigação, as suas principais características, bem como as consequências negativas da sua utilização. Enumeraremos seguidamente algumas das iniciativas para a reforma da avaliação da investigação e tentaremos identificar os pontos comuns que podem constituir o núcleo da reforma. Por fim procuraremos apresentar e discutir brevemente os principais desafios e obstáculos para a mudança da avaliação de investigação, que fazem desta indispensável reforma uma tarefa difícil.

2 - As origens do atual modelo de avaliação

O modelo de avaliação da investigação que hoje predomina, baseado em indicadores bibliométricos, foi estabelecido nas últimas décadas, e está intimamente associado à evolução do sistema de comunicação científica e académica que ocorreu desde meados do século passado. Por isso, vale a pena determo-nos brevemente sobre o que aconteceu na comunicação científica, antes de abordarmos a avaliação da investigação.

A publicação científica, como hoje a conhecemos, tem as suas raízes na segunda metade do século XVII com a criação das primeiras revistas científicas¹. As revistas, como a *Philosophical Transactions* que ainda hoje se publica, criadas a partir de então contribuíram para afirmar a dimensão internacional da ciência, estabelecer os procedimentos de controlo

¹ As duas primeiras foram criadas no ano de 1665: o *Journal des sçavans*, em 5 de janeiro, e a *Philosophical Transactions of the Royal Society*, em 6 de março.

de qualidade e revisão pelos pares e os ideais de objetividade que, no seu conjunto, são característicos do empreendimento científico (Johns, 2000).

Como referimos em outra ocasião (Rodrigues, 2020)², até meados do século XX a publicação académica manteve-se relativamente estável, como uma atividade não lucrativa e fiel ao seu objetivo central e original de disseminar e partilhar os resultados da investigação. A publicação de revistas (e da maioria dos livros académicos) era realizada por sociedades científicas ou editoras universitárias que suportavam e subsidiavam os custos da circulação do conhecimento na comunidade académica (Fyfe et al., 2017).

No entanto, após a segunda guerra mundial, a situação alterou-se. O acentuado crescimento da atividade científica, suportado por um forte investimento dos estados na investigação (nomeadamente das potências vencedoras, que mantiveram uma “corrida” científica e tecnológica nas décadas seguintes) e a expansão do ensino universitário, criaram as condições para uma mudança profunda no sistema de comunicação científica e académica. Esse processo de transformação iniciou-se nos anos de 1950, tendo-se alargado e consolidado nas décadas seguintes, e encontrava-se no essencial concluído no final do século passado.

A mudança passou pelo estabelecimento de novas estratégias de publicação, com orientação comercial, por parte de algumas empresas (como a Elsevier ou a Pergamon Press), que souberam aproveitar o crescimento da oferta e da procura, para criar uma indústria altamente lucrativa. O sucesso dessa estratégia fez com que ela fosse imitada por muitas empresas, mas também por editoras académicas e não comerciais (Fyfe et al., 2017) e resultou ainda na integração de muitas revistas, até então publicadas por sociedades científicas, no portfólio de publicação das editoras comerciais.

Finalmente, desde os finais da década de 1960 que se assistiu a um intenso movimento de aquisições e fusões que resultou no atual mercado editorial altamente concentrado e com características de oligopólio. Apesar do crescimento do número de revistas e artigos publicados, estima-se que os cinco principais grupos editoriais sejam responsáveis por mais de 50% de todos os artigos publicados (Larivière et al., 2015).

² Retomamos e sintetizamos aqui o que sobre as origens e evolução do sistema de comunicação académica escrevemos com mais detalhe em Rodrigues, E. (2020). A pandemia e a emergência da ciência aberta. In M. Martins & E. Rodrigues (Eds.), *A Universidade do Minho em tempos de pandemia: Tomo II: Re(Ações)* (Vol. 2, pp. 263–294). UMinho Editora.
<https://doi.org/10.21814/uminho.ed.24.12>

O mercado da informação científica gera receitas anuais de mais de vinte e cinco mil milhões de dólares, dos quais cerca de dez mil milhões são provenientes da publicação de revistas científicas (os restantes são de bases de dados e ferramentas de software, livros, informação técnica, etc.) (STM, 2021). Os grandes grupos editoriais que dominam este mercado, apresentam elevada rentabilidade, com taxas de lucro em torno ou acima dos 30%³.

A mudança no modelo e no mercado editorial e o estabelecimento do atual modelo de avaliação da investigação foram simultâneas. Apesar de coincidentes no tempo, os dois fenómenos começaram por ser, aparentemente, independentes. Mas, pelo menos desde a década de 1990, as duas dimensões (mercado editorial e avaliação académica) evoluíram em estreita relação e influenciando-se mutuamente, criando uma espécie de “economia de prestígio”, com as métricas a contribuir para o reforço da posição dominante de algumas revistas e grupos editoriais, e a servir como justificação para a escalada dos preços.

A publicação científica é, desde há muito, um elemento relevante dos sistemas, formais ou informais, de avaliação, reconhecimento e recompensa na comunidade académica. Ao longo do século XX, a publicação foi ocupando um lugar cada vez mais importante na cultura e nas práticas dos investigadores universitários e na avaliação do seu desempenho individual, e da qualidade institucional das organizações onde trabalham. Mas foi sobretudo nas últimas décadas do século XX que a avaliação baseada em métricas se foi tornando dominante.

As mudanças na avaliação da investigação foram o produto da convergência de dois fatores. Por um lado, uma crescente exigência de responsabilização e prestação de contas (*accountability*) por parte dos governos e entidades financiadoras relativamente às instituições que financiam, e, conseqüentemente, destas relativamente ao seu pessoal docente e investigador. Esta exigência e necessidade de prestação de contas às autoridades e à sociedade, que acompanhou o crescimento do investimento realizado no ensino superior, na ciência e tecnologia, requereu a identificação, produção e disponibilização de indicadores que demonstrassem os resultados e o impacto desses investimentos. Na década de 1980, a utilização de indicadores mensuráveis do valor da investigação ocorria sobretudo no Reino Unido e nos EUA, mas na década seguinte a utilização de indicadores bibliométricos começou a generalizar-se (Wilsdon et al., 2015).

³ Por exemplo, a Elsevier reportou um volume de receitas de 2,5 mil milhões de libras e um lucro de 942 milhões de libras em 2018, aumentando a sua margem de lucro de 37% (em 2017) para 37,1% (Matthews, 2019).

Por outro lado, as mudanças só foram possíveis porque, a partir da década de 1960, começaram a estar disponíveis bases de dados e ferramentas cientométricas e bibliométricas. O campo da cientometria desenvolveu-se ao longo do século XX⁴, mas foram as ferramentas bibliométricas e cientométricas inventadas por Eugene Garfield nas décadas de 1950 e 1960, como o *Science Citations Index* (Garfield, 1955) e os *Journal Citation Reports* (em 1975) de que resultou o *Journal Impact Factor*, ou JIF⁵ (Garfield, 2006), que revolucionaram a área.

O *Science Citations Index*, criado em 1961, foi inicialmente utilizado sobretudo pelos especialistas em cientometria, para descrever o desenvolvimento e os rumos da investigação científica, e não para avaliar a sua qualidade (Wilsdon et al., 2015). Até ao final da década de 1980, a utilização dos dados bibliométricos e das métricas, como o JIF, na avaliação era pouco frequente e estava concentrada no mundo anglo-saxónico. Mas, a partir da década de 1990, o uso de métricas na avaliação generalizou-se.

Finalmente, desde o princípio do século XXI, assistiu-se a um significativo aumento da acessibilidade de ferramentas bibliométricas, quer pela disponibilização na World Wide Web das já anteriormente existentes (como o *Science Citation Index* e outras ferramentas criadas por Garfield, disponibilizadas como *Web of Science* a partir de 2001), quer pela criação de novas ferramentas (por exemplo, a *Scopus*, da *Elsevier*, em 2003, ou o *Google Scholar* em 2004) e métricas de citação como o h-index (índice h ou índice de Hirsch), o *Eigenfactor Score*, o *Article Influence Score* (AIS), ou as designadas métricas alternativas (*altmetrics*), que procuram representar outros tipos de visibilidade e impacto, como as citações/links na Web, ou referências nas redes sociais.

3 – Limitações e efeitos da avaliação baseada em métricas

A avaliação da investigação e dos investigadores excessivamente (quando não exclusivamente) baseada em métricas de publicação e citação, e frequentemente em métricas indiretas como o fator de impacto das revistas, que se tornou dominante nos últimos 30 anos, tem revelado muitos problemas de qualidade e fiabilidade, e teve um impacto muito significativo (e em muitos aspetos negativo) no sistema académico e científico. As limitações

⁴ Já em 1926, Alfred J. Lotka publicou o artigo sobre a que se viria a designar “Lei de Lotka”, que descreve a frequência de publicação de autores em qualquer área, revelando que um baixo número de autores é responsável por uma grande percentagem das publicações numa determinada área.

⁵ O JIF foi criado por Garfield como um indicador da relevância de cada revista científica no seu domínio, refletindo o número médio anual de citações que os artigos publicados nos últimos dois anos numa determinada revista receberam.

e consequências da avaliação estão profundamente entrelaçadas, mas, por conveniência de exposição, iremos apresentá-las sequencialmente.

Em primeiro lugar, assinale-se o facto da avaliação se concentrar na publicação (e no seu impacto associado ao número de citações), ignorando quase sempre todas as outras atividades e contributos (como a geração, curadoria e processamento de dados de investigação, ou a revisão por pares do trabalho de terceiros, para dar apenas dois exemplos) que são essenciais para a investigação, e sem as quais a publicação não existiria. Como já foi assinalado⁶, a avaliação considera apenas a pequena ponta visível do icebergue que é a publicação final, desprezando a maior parte do trabalho científico que lhe dá origem.

Ainda mais grave é que, frequentemente, o foco da avaliação não é sequer a publicação (o seu conteúdo), mas o local da publicação (ou seja, o contentor). Na maioria das instituições de investigação, como as universidades, é usado o fator de impacto, ou indicadores a ele associados como os quartis do JCR, nos processos de avaliação⁷. Isto apesar de se saber que o JIF é uma métrica criada para avaliar a “qualidade” média de revistas (e não de pessoas ou grupos de investigação) e se reconhecer ser um indicador muito grosseiro. Como o fator de impacto é uma média de citações, não representa a distribuição das citações, que geralmente apresenta uma grande variação. Por exemplo, nas revistas de elevado fator de impacto, como a *Nature* ou a *Science*, uma pequena percentagem (cerca de 15%) dos artigos é responsável por metade das citações (ou seja, do fator de impacto), enquanto a maioria (cerca de 75%) dos artigos recebe um número de citações inferior ao fator de impacto da revista (Larivière et al., 2016), incluindo cerca de 10% dos artigos publicados que recebem 0 ou 1 citações.

No entanto, nos processos de avaliação baseados no JIF ou nos quartis das revistas, um investigador que publique um artigo na *Nature* com 0 citações terá, por esse artigo, a mesma avaliação que outro que publique na mesma revista um dos artigos com centenas de citações, e melhor avaliação que o seu colega que tenha publicado um artigo com dezenas de citações, mas numa revista com um fator de impacto inferior.

⁶ Ver a ilustração no post Roskams-Edris, D., Poline, JB., Stikov, N. (2019, Agosto 15). The Canadian Open Neuroscience Platform: Catching Up to Plan S and Going Further. *PLoS Blogs* . <https://theplosblog.plos.org/2019/08/the-canadian-open-neuroscience-platform-catching-up-to-plan-s-and-going-further/>, consultado em 3 de junho de 2022.

⁷ Cerca de 75% das universidades europeias, que responderam ao inquérito da Associação Europeia de Universidades (EUA), utilizavam o Journal Impact Factor nos seus processos de avaliação do pessoal académico e investigador (Saenen et al., 2019)

Medir a qualidade de um artigo através do fator de impacto da revista onde foi publicado, recorda a expressão popular portuguesa sobre a pescada que antes de ser já o era! Quando se baseia no fator de impacto, a avaliação já não ocorre após a publicação, mas sim em conjunto com o ato de publicação (Biagioli & Lippman, 2020b). E como assinalam estes autores, ao passar a representar o impacto/valor de cada artigo individual o fator de impacto

(...) functions, literally, like money. It is the “face value” impressed on a coin or banknote, which determines its value no matter what the coin is made of (copper, silver, gold, steel), or what the exchange value that coin would have as metal. The “content” of the coin is just the medium for the stamp, which needs to be impressed on some material. What carries value is not the inside but the outside—the number inscribed on the surface. Whether the article is a piece of gold or lead (or worse), the JIF has come to determine the face value of that article.

(...)

Attaching an impact factor to an article at the time of publication (and thus before real impact has accrued) shows that the JIF has literally nothing to do with the evaluation of that specific article, but simply prices it in a currency that allows for exchange. As an author, I can “trade” articles with certain impact factors into a job, and the institution that employs me can then “trade” those publications (together with hundreds or thousand more by other faculty of the same university) into a better national or global ranking, which may be subsequently traded into more students, donors, contracts, and so on.” (Biagioli & Lippman, 2020b)

Finalmente, no que diz respeito às limitações da avaliação baseada nas métricas de publicação, importa assinalar que o uso das citações como indicador de impacto, qualidade e relevância, não está isento de problemas e riscos, que são quase sempre ignorados nos atuais sistemas de avaliação. Para além das más práticas ou comportamentos fraudulentos, de que falaremos seguidamente, que influenciam o número de citações, os fatores de impacto das revistas ou o h-index dos investigadores, também a natureza da citação (“positiva” de reconhecimento, ou “negativa” para contrariar), bem como as diferenças na cultura e nas práticas das diferentes disciplinas científicas na autoria (nomeadamente no número de co-autores) e na citação não são devidamente consideradas nos sistemas de avaliação baseados em métricas simples.

Em grande medida pelas limitações aqui assinaladas, a avaliação baseada em métricas, produziu diversos efeitos, muitos dos quais com impacto e consequências negativas para a investigação e o sistema científico.

Isso resulta, e como foi previsto logo nos anos de 1970 pelo sociólogo Robert Merton, do designado deslocamento do objetivo (*goal displacement*). Em 1974, em diversas interações com Eugene Garfield, Merton fez uma previsão que se revelou acertada, e antecipou a que viria a ficar conhecida como lei de Goodhart⁸ no domínio económico:

“In the case of citations, I make a small self-fulfilling prophecy, that has already been fulfilled in one area I know, namely, as more and more citations are used both officially and unofficially as measures of contribution of work of relative standing and the like, there will be diverse patterns of adaptations to those uses, in a feedback way, and citation behavior will change in part in certain sub-sets of the community.”

“Watch out for goal displacement: Whenever an indicator comes to be used in the reward system of an organization or institutional domain, there develop tendencies to manipulate the indicator so that it no longer indicates what it once did. Now that more and more scientists and scholars are becoming acutely conscious of citations as a form of behavior, some will begin, in semi-organized fashion, to work out citation-arrangements in their own little specialties.” (Robert Merton, citado em Csiszar, 2020)

De facto, nas últimas décadas observou-se uma profunda mudança de comportamentos na comunidade científica no sentido da adaptação aos sistemas de avaliação métricos. Desde logo, a natural tendência para não apenas publicar mais, mas para publicar nas revistas de maior fator de impacto possível. O mantra “publicar, ou perecer” foi acrescentado para se tornar “publicar em revistas de alto fator de impacto, ou perecer”.

Poder-se-á considerar que esta pressão para o aumento da publicação científica e do maior impacto possível não é negativa em si mesma. Mas os efeitos que ela produz levam-nos a questionar essa visão benigna. O condicionamento da agenda de investigação, ou seja, a escolha dos temas/objetos e das metodologias utilizadas em função do seu potencial de publicação e não da sua relevância, é um dos efeitos dessa pressão. No limite, a relação entre investigação e publicação pode ficar invertida: não se publica porque se investiga, mas investiga-se de forma a publicar.

O atraso na divulgação de resultados, não só pela demora no processo de revisão por pares (que também é, em parte, consequência desta corrida à publicação), mas igualmente pelas sucessivas tentativas de publicação (os autores tentam publicar na revista de maior fator de

⁸ Formulada posteriormente como “When a measure becomes a target, it ceases to be a good measure”.

impacto, e em face de rejeição vão descendo a escada dos fatores de impacto até encontrar uma revista que aceite a publicação), é outra consequência desta cultura.

Um terceiro efeito é a pressão para a convergência e uniformização das práticas de publicação e citação em todas as áreas científicas, começando a esbater-se as diferenças das ciências humanas e sociais (nas quais, em diversas disciplinas, o artigo não é o artefacto principal para a comunicação, mas sim o livro ou o capítulo de livro) face às restantes ciências e engenharias. Por exemplo, a tendência para o aumento do número de autores em cada publicação (Aboukhalil, 2015) verifica-se também nas ciências sociais (Henriksen, 2016). E se o aumento do número médio de autores terá várias razões objetivas que o explicam (como o aumento das colaborações na investigação, que acompanha o crescimento da dimensão e complexidade dos projetos de investigação), ele não pode ser dissociado da pressão para publicar e do crescimento de práticas questionáveis, ou mesmo más práticas, na atribuição de autoria.

O crescimento das más práticas na investigação e publicação científica é, aliás, um dos efeitos mais negativos da cultura métrica que tem dominado a comunidade e as instituições científicas nas últimas décadas. As más práticas, como a fabricação e falsificação de resultados ou o plágio, sempre terão existido mas ter-se-ão tornado mais frequentes nas últimas décadas⁹, sendo também uma das principais razões para a retirada de artigos científicos (Fang et al., 2012), fenómeno que tem também crescido (Steen et al., 2013)¹⁰, existindo muitos artigos retirados que receberam um número significativo de citações, incluindo citações após a sua retirada (Hsiao & Schneider, 2021) (Oransky, 2022).

E foi surpreendente a revelação da existência de correlação entre o fator de impacto e a frequência da retirada de artigos: ou seja, a frequência de artigos retirados é mais elevada nas revistas com maior fator de impacto (Fang e Casadevall, 2011; Why High-Profile Journals Have More Retractions, 2014). Essa relação, contraintuitiva, entre o aumento dos problemas de controlo de qualidade e as revistas de maior impacto e prestígio, foi verificada também

⁹ Com alguns casos muito mediáticos como a “clonagem humana” que Woo Suk Hwang anunciou em dois artigos na *Science* em 2004 e 2005, com base em dados falsificados (Cyranoski, 2006). No momento em que terminamos a redação deste texto, a revista *Science* noticiou que uma das principais teorias e linhas de investigação de Alzheimer está baseada em imagens manipuladas/fabricadas (Piller, 2022)

¹⁰ Também em Portugal. De acordo com a base de dados Retraction Watch, consultada em julho de 2022, foram retirados 15 artigos com autores de Portugal publicados entre 2000 e 2010, mas esse número sobe para 89 relativamente ao período de 2011 a 2020. Note-se que é expectável que este último número venha a subir dado que muitos artigos são retirados 2, 3 ou mais anos após a sua publicação.

para publicações que não foram retiradas. Foram já apresentadas evidências que a confiabilidade dos resultados e a sua reprodutibilidade diminui com o aumento do fator de impacto das revistas (Brembs, Button e Munafò, 2013; Brembs, 2018).

As causas da já designada crise de reprodutibilidade (com taxas de irreprodutibilidade superiores a 50%, e nalguns casos a 70%, em várias disciplinas) são múltiplas (Miyakawa, 2020). Mas é legítimo considerar que em várias delas (desde más práticas na gestão de dados, até ao reporte seletivo de resultados positivos) existe uma relação com a avaliação baseada nas métricas de publicação.

Para além das más práticas que afetam o seu conteúdo, ocorrem outros problemas em torno das publicações que podem também ser associados à cultura de “publicar o mais possível, em revistas com maior impacto possível”. Práticas questionáveis de autoria, falsas autorias, trocas de autoria ou “autorias de favor” (“eu coloco-te como autor neste meu artigo e, em troca, tu fazes o mesmo no teu artigo”) são um exemplo deste tipo de problemas. Também a manipulação ou falsificação da revisão por pares já foi detetada posteriormente em centenas de ocasiões, originando a retirada de artigos (Ferguson et al., 2014; Oransky, 2020).

Mas os fenómenos mais significativos, e literalmente mais impactantes (pois refletem-se nos fatores de impacto e outros indicadores), são as várias formas de manipulação de citações. O uso excessivo de auto-citações, os círculos de citação entre autores (para aumentar o número de citações de cada um), ou entre revistas da mesma editora (para melhorar os seus fatores de impacto), o (ab)uso do processo editorial e de revisão por pares para “sugerir” ou “encomendar” citações (dos próprios revisores ou dos seus grupos de investigação, de outros artigos da mesma revista ou grupo editorial) são fenómenos conhecidos e repetidamente reportados (Wilhite & Fong, 2012; Van Noorden, 2012, 2020a, 2020b; Van Noorden & Singh Chawla, 2019; Singh Chawla, 2019). Tendo consciência do problema o Committee on Publication Ethics elaborou um documento com algumas orientações (COPE Discussion Document, 2019) e diversas revistas, como os *Proceedings of the Royal Society A* (Lockwood, 2020) tem tomado medidas para a deteção e mitigação das más práticas, combinando a melhoria dos procedimentos editoriais com o uso de algoritmos.

Uma das características destas manipulações em torno das publicações é que estas não se restringem aos indivíduos (como geralmente acontece com as más práticas relacionadas com o conteúdo), mas envolvem grupos, redes ou mesmo instituições (Biagioli & Lippman, 2020b). A crescente importância e utilização dos rankings internacionais de instituições tem feito escalar as práticas questionáveis e a manipulação de dados também para o nível das

universidades (Kehm, 2020).

4 - A reforma do sistema de avaliação da investigação

A crescente evidência e reconhecimento das limitações e das consequências adversas do modelo de avaliação baseado em métricas de publicação e citação fez emergir, na última década, diversas iniciativas para a sua reforma. A primeira, e mais conhecida, foi a *San Francisco Declaration on Research Assessment* ou DORA¹¹, publicada em 2013, recomendando o abandono de métricas baseadas nas revistas, como o fator de impacto, como indicador da qualidade dos artigos individuais e na avaliação dos investigadores para contratação, promoção ou financiamento. A DORA transformou-se, em 2018, numa iniciativa internacional para apoiar e facilitar a reforma da avaliação da investigação.

Outra declaração que obteve elevada visibilidade foi o *Leiden Manifesto for research metrics*¹² (Hicks et al., 2015), em 2015, cujos dez princípios começam por enfatizar que a avaliação quantitativa deve apoiar a avaliação qualitativa especializada, recomendando também que a avaliação dos investigadores individuais se deve basear na apreciação qualitativa do seu portfólio.

Após estes dois documentos seminais, têm sido publicados na Europa vários estudos e recomendações, de entre os quais destacamos o *The Metric Tide* (Wilsdon et al., 2015), e o relatório do *Working Group on Rewards under Open Science* da Comissão Europeia (O'Carroll et al., 2017), onde se recomenda a utilização de uma matriz de indicadores, e não métricas, na avaliação.

Sobretudo a partir de 2015, as iniciativas relativas à reforma da avaliação da investigação passaram a estar relacionadas e influenciadas pela consolidação dos movimentos de acesso aberto e ciência aberta. As duas agendas, a da reforma da avaliação e a da ciência aberta, possuem vários pontos comuns e são hoje significativamente interdependentes, pois considera-se que não são possíveis progressos significativos em qualquer uma delas, sem que o mesmo aconteça na outra. Exemplos desta convergência são os documentos e iniciativas da *EUA - European University Association* neste domínio¹³, bem como o Seminário

¹¹ <https://sfdora.org/>

¹² <http://www.leidenmanifesto.org/>

¹³ Ver, entre outros, (European University Association, 2018; Saenen et al., 2019, 2021; Saenen & Borrell-Damián, 2019), bem como a série de webinars organizada em 2020

promovido pelos Conselhos de Reitores de Portugal (CRUP) e Espanha (CRUE) em 2020¹⁴, que cruzam os dois temas.

Para além dos estudos, recomendações e debates, no final da década passada começaram a ocorrer as primeiras iniciativas de implementação prática de reformas da avaliação na Europa, quer individualmente em algumas universidades, quer a nível nacional. Um dos primeiros casos conhecidos é o da Universidade de Ghent, que em 2016 aprovou e publicou uma Declaração (*Vision Statement*) para avaliação da investigação, com base em oito princípios acordados pelo *Board of Governors* da Universidade¹⁵, e em 2018 introduziu três profundas mudanças na avaliação do seu pessoal académico: a passagem dos períodos de avaliação para cinco anos, a aposta na orientação profissional (*coaching*) com a atribuição de uma equipa de orientação a cada professor e, finalmente, a avaliação baseada no relatório de cada avaliado e não nas suas métricas de publicação¹⁶.

Várias outras universidades, em diversos países europeus, iniciaram nos últimos anos processos de revisão dos seus procedimentos de avaliação, diminuindo o papel das métricas de publicação (Saenen et al., 2021). Mas, para além das iniciativas individuais de universidades, têm vindo a desenvolver-se também iniciativas a nível nacional, envolvendo instituições de investigação, agências de financiamento e governos. Esse é o caso da Holanda, onde as principais organizações ligadas à investigação, como a VSNU (Associação de Universidades Holandesas), a KNAW (Academia Real das Artes e Ciências) e o NWO (o Conselho de Investigação que é um financiador de investigação) produziram um documento orientador (*Room for everyone's talent - towards a new balance in the recognition and rewards for academics*, 2019), cujas recomendações estão a ser implementadas nas universidades holandesas, e estão também incorporadas no *Strategy Evaluation Protocol* (SEP) (VSNU, KNAW and NWO, 2020), que descreve os objetivos e métodos utilizados para avaliar a investigação nas universidades holandesas, bem como nos institutos do NWO e KNAW.

Na Noruega, a associação de universidades desenvolveu uma matriz de avaliação de carreiras (NOR-CAM - *The Norwegian Career Assessment Matrix*), baseada em seis princípios e quatro recomendações, que pode ser usada não apenas pelas instituições académicas,

¹⁴ O CRUP e a CRUE realizaram conjuntamente, em junho de 2020, o 2nd CRUP/CRUE Open Science Seminar: Career Assessment in the transition to Open Science - <https://www.youtube.com/playlist?list=PLIm8KTWNUm7XF4WLtFMmylvvm2OMKHLg>

¹⁵ <https://www.ugent.be/en/research/research-strategy/evaluation/principles.htm>

¹⁶ Para mais informação sobre o atual processo de avaliação da Universidade de Ghent ver <https://www.ugent.be/en/work/mobility-career/career-aspects/professorial-staff> e (Redden, 2019).

mas também pelas agências de financiamento e pelas autoridades nacionais (Universities Norway, 2021).

Em outros países, como a França (Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, 2021) e a Irlanda (Government of Ireland - National Open Research Forum, 2019), os planos ou documentos de estratégia para a ciência aberta incluem também recomendações ou medidas concretas relativas à reforma do sistema de avaliação da investigação. O mesmo se teria passado em Portugal, se as recomendações do Grupo de Trabalho para a Política Nacional de Ciência Aberta (Comissão Executiva GT-PNCA, 2016), que incluíam diversas recomendações sobre a avaliação da investigação, tivessem sido aprovadas e implementadas pelo governo.

Ainda que a maioria das iniciativas de reforma da avaliação sejam europeias¹⁷, elas ocorrem também em várias outras geografias. É o caso da China, que anunciou em 2020 uma reforma no seu sistema de avaliação (China's Research-Evaluation Revamp Should Not Mean Fewer International Collaborations, 2020; Zhang & Sivertsen, 2020), abandonando o que designaram por “adoração do *Science Citation Index*” que se verificou nas últimas décadas, e deixando de atribuir prémios monetários aos investigadores pela publicação em revistas do SCI (Mallapaty, 2020). Na América Latina não são conhecidas iniciativas práticas de reforma, mas foi constituído, em 2019, o *Foro Latinoamericano sobre Evaluación Científica* (FOLEC) para o debate e partilha de experiências e alinhamento de práticas e ferramentas a nível regional. Por último, a nível global, na Recomendação da UNESCO sobre Ciência Aberta (UNESCO, 2021), existem várias menções à necessidade de reformar a avaliação da investigação e investigadores, alinhando-a com a ciência aberta.

Apesar da grande diversidade e das especificidades de cada uma das declarações, documentos e iniciativas de reforma da avaliação aqui referidas, existem três pontos principais que estão presentes, de uma ou outra forma, em quase todas. O primeiro diz respeito ao abandono do uso exclusivo de métricas de publicação, em particular métricas indiretas como o *Journal Impact Factor*, na avaliação. E à prescrição de um uso responsável de métricas recomendando-se, em muitos casos, que as métricas sejam usadas apenas como complemento de avaliações qualitativas, que devem constituir o centro da avaliação.

¹⁷ Como a mais recente e relevante iniciativa europeia lançada pela Comissão Europeia, e que já reuniu mais de 350 organizações de 40 países, e cujo texto final (Agreement on Reforming Research Assessment, 2022) foi anunciado a 20 de julho de 2022, que mencionaremos mais abaixo, e é mais detalhadamente referida noutro capítulo deste livro (Ver “Publicação Científica, Ciência Aberta e Avaliação de Investigação – Breve Análise Crítica”)

O segundo refere-se à necessidade de ter uma abordagem mais abrangente e flexível da avaliação, tomando em consideração, incentivando e recompensando todos os contributos relevantes numa gama mais vasta de atividades académicas, não apenas relativas à investigação, mas também ao ensino, interação com a sociedade, ou gestão institucional.

Finalmente, o terceiro, é relativo ao acolhimento e reconhecimento da diversidade dos contextos (disciplinares, institucionais, nacionais), das carreiras e da fase em que os investigadores e outro pessoal académico se encontram nessas carreiras, nos processos de avaliação, nomeadamente na seleção dos indicadores e no respetivo peso.

Estes três pontos recebem também acolhimento no recente documento do Conselho da União Europeia sobre a avaliação da investigação e a implementação da ciência aberta (Council of the European Union, 2022).

5 – O difícil caminho para a reforma da avaliação da investigação

Existindo, há cerca de uma década, um generalizado reconhecimento dos defeitos e consequências negativas da utilização do fator de impacto na avaliação da investigação, porque é que ele continua a ser amplamente utilizado pela maioria das universidades (Tonta, 2014; Else, 2019; McKiernan, et al., 2019; Saenen et al., 2019)? Existindo, há vários anos, um amplo consenso, não apenas sobre a necessidade de reformar os sistemas de avaliação, mas também sobre os contornos gerais dessa reforma, porque se tem revelado lenta e difícil a sua concretização?

As razões serão certamente múltiplas, mas identificamos aqui brevemente três das que consideramos mais relevantes. Em primeiro lugar, o que se pode designar por inércia, tradição ou conservadorismo académico. Apesar de todas as críticas e evidências das suas limitações, as noções de qualidade, excelência e prestígio associadas ao fator de impacto e às “marcas” das revistas, das editoras e dos “índices” (como o Science Citation Index) estão profundamente enraizadas na comunidade científica e na cultura académica, quer ao nível individual (Grove, 2022), quer ao nível institucional. Por outro lado, sobretudo no pessoal académico com carreiras com mais de uma década, desenvolvidas no quadro do sistema métrico de avaliação que moldou o seu comportamento e as suas expectativas, é natural a existência de alguma inércia, ou mesmo resistência e oposição, à mudança de paradigma na avaliação.

Em segundo lugar, estão as questões relacionadas com a facilidade/dificuldade e o nível de esforço requerido nos processos de avaliação. Os processos baseados na utilização de métricas, como o fator de impacto (ou quartis) das revistas ou o h-index, são tendencialmente simples e requerem pouco esforço. Dispondo dos números e de folhas de cálculo, os avaliadores terão a sua tarefa facilitada e poderão concluí-la em pouco tempo. Pelo contrário, as avaliações qualitativas, que requeiram a leitura de publicações ou outros documentos apresentados pelos avaliados, exigem não apenas mais tempo, mas igualmente um nível superior de ponderação, compromisso ético e responsabilização por parte dos avaliadores. No contexto atual, no qual muito pessoal académico se queixa de demasiado trabalho e falta de tempo, este é um obstáculo importante para o abandono da avaliação essencialmente quantitativa. E é também um problema para o qual as instituições, como as universidades, terão de encontrar soluções que não se afiguram fáceis.

Finalmente, e em terceiro lugar, queremos referir o ambiente de hiper-competitividade que se instalou no mundo académico e científico nas últimas décadas. Com alguma variabilidade no grau e dimensão, existe hoje na maioria dos países e instituições uma grande competição por recursos que são escassos: pelo financiamento, por posições/carreira académica, por estudantes. Quer ao nível individual, lutando pela progressão na carreira e pela obtenção de financiamento dos seus projetos, quer ao nível institucional, procurando captar financiamento, atrair muitos e bons estudantes e ficar bem posicionado nos rankings, poderá existir algum receio e ansiedade em abandonar em primeiro lugar, ou sozinho, os comportamentos e as ferramentas que são recompensados pelos processos de avaliação atuais.

Já foram reportadas consequências negativas da hiper-competitividade na saúde mental dos investigadores (Wellcome Trust, 2020). A nível institucional, o ambiente ultracompetitivo no domínio da investigação e do ensino, origina muitas pressões e obstáculos, internos e externos, para a mudança da cultura organizacional e dos processos de avaliação. Uma das preocupações centrais será o impacto que as mudanças na avaliação, no sistema de incentivos e recompensas e, conseqüentemente, no desempenho dos seus membros e nos resultados obtidos (número de publicações, financiamento captado, etc.) pode ter na sua posição nos diversos rankings e na sua reputação que se associa a esses rankings¹⁸ (Hatch & Curry, 2020).

¹⁸ Isto apesar de, à semelhança do que se passa com as métricas de publicação (que, aliás constituem uma das componentes centrais dos rankings), se questionar cada vez mais a qualidade, fiabilidade e relevância dos rankings como indicadores da qualidade institucional. Já tem sido notado que os dados dos rankings não são geralmente abertos, nem sujeitos a escrutínio e revisão por pares (Wilsdon et

As universidades que sejam precoces, ou que introduzam mudanças mais profundas na melhoria das suas práticas de avaliação podem sentir-se, e estar objetivamente, a correr alguns riscos. Mas, admitindo como provável que a reforma da avaliação vai mesmo acontecer nos próximos anos, serão também essas instituições pioneiras as primeiras a beneficiar de se encontrarem ajustadas e preparadas para os novos paradigmas de avaliação da investigação.

Em qualquer caso, parece claro que a reforma da avaliação da investigação, a nível institucional, se fará tanto melhor e mais rapidamente quanto seja partilhada e assumida coletivamente por um número significativo de instituições. Por isso, a recente iniciativa europeia (Agreement on Reforming Research Assessment, 2022) assume uma grande importância. Desde logo, porque já reuniu, na fase de preparação, um número muito significativo de instituições, em particular universidades e agências de financiamento, que constituirão a base da Coligação que se irá formalmente estabelecer no outono de 2022.

E, o que é ainda mais relevante, porque as organizações signatárias se comprometem não apenas com a aplicação calendarizada de práticas de avaliação que reconheçam e valorizem a diversidade e abandonem o uso inadequado de métricas, como assumem também o compromisso de tornar públicos os seus planos e progressos, e partilhar práticas e experiências, para permitir a aprendizagem mútua dentro e fora da Coligação.

Referências bibliográficas

Aboukhalil, R. (2015). The rising trend in authorship. *The Winnower*.

<https://doi.org/10.15200/winn.141832.26907>

Agreement on Reforming research assessment. (2022).

https://eua.eu/downloads/news/2022_07_19_rra_agreement_final.pdf

Biagioli, M., & Lippman, A. (Eds.). (2020a). *Gaming the Metrics: Misconduct and Manipulation in Academic Research*. The MIT Press. <https://doi.org/10.7551/mitpress/11087.001.0001>

al., 2015). Existem também conflitos de interesse inaceitáveis, como a prestação de serviços de consultoria para melhorar a posição nos rankings, por parte de empresas que os produzem.

- Biagioli, M., & Lippman, A. (2020b). Introduction: Metrics and the New Ecologies of Academic Misconduct. In *Gaming the Metrics: Misconduct and Manipulation in Academic Research* (pp. 1–23). The MIT Press. <https://doi.org/10.7551/mitpress/11087.003.0001>
- Brembs, B. (2018). Prestigious Science Journals Struggle to Reach Even Average Reliability. *Frontiers in Human Neuroscience*, 12, 37. <https://doi.org/10.3389/fnhum.2018.00037>
- Brembs, B., Button, K., & Munafò, M. (2013). Deep impact: Unintended consequences of journal rank. *Frontiers in Human Neuroscience*, 7. <https://doi.org/10.3389/fnhum.2013.00291>
- China's research-evaluation revamp should not mean fewer international collaborations. (2020). *Nature*, 579(7797), 8–8. <https://doi.org/10.1038/d41586-020-00625-0>
- Comissão Executiva GT-PNCA. (2016). *Segundo relatório da Comissão Executiva do Grupo de Trabalho para a Política Nacional de Ciência Aberta (GT-PNCA)*. https://www.ciencia-aberta.pt/files/ugd/a8bd7c_3274046fc8ce42c78db2ec1707c0a0fd.pdf
- Committee on Publication Ethics. (2019). *COPE Discussion Document: Citation Manipulation*. <https://doi.org/10.24318/cope.2019.3.1>
- Council of the European Union. (2022). *Research assessment and implementation of Open Science—Council conclusions (adopted on 10 June 2022)*. <https://data.consilium.europa.eu/doc/document/ST-10126-2022-INIT/en/pdf>
- Csiszar, A. (2020). Gaming Metrics Before the Game: Citation and the Bureaucratic Virtuoso. In M. Biagioli & A. Lippman (Eds.), *Gaming the Metrics: Misconduct and Manipulation in Academic Research* (pp. 31–42). The MIT Press. <https://doi.org/10.7551/mitpress/11087.003.0003>
- Cyranoski, D. (2006). Verdict: Hwang's human stem cells were all fakes. *Nature*, 439(7073), 122–122. <https://doi.org/10.1038/439122a>
- Else, H. (2019). Impact factors are still widely used in academic evaluations. *Nature*. <https://doi.org/10.1038/d41586-019-01151-4>
- European University Association. (2018). *EUA Roadmap on Research Assessment in the Transition to Open Science*. https://eua.eu/downloads/publications/eua-roadmap-on-research-assessment-in-the-transition-to-open-science_v20-08-2019.pdf
- Fang, F. C., & Casadevall, A. (2011). Retracted Science and the Retraction Index. *Infection and Immunity*, 79(10), 3855–3859. <https://doi.org/10.1128/IAI.05661-11>
- Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences*, 109(42), 17028–17033. <https://doi.org/10.1073/pnas.1212247109>
- Ferguson, C., Marcus, A., & Oransky, I. (2014). Publishing: The peer-review scam. *Nature*, 515(7528), 480–482. <https://doi.org/10.1038/515480a>

- Fyfe, A., Coate, K., Curry, S., Lawson, S., Moxham, N., & Røstvik, C. M. (2017). *Untangling Academic Publishing: A history of the relationship between commercial interests, academic prestige and the circulation of research*. Zenodo. <https://doi.org/10.5281/ZENODO.546100>
- Garfield, E. (1955). Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. *Science*, 122(3159), 108–111. <https://doi.org/10.1126/science.122.3159.108>
- Garfield, E. (2006). The History and Meaning of the Journal Impact Factor. *JAMA*, 295(1), 90. <https://doi.org/10.1001/jama.295.1.90>
- Government of Ireland - National Open Research Forum. (2019). *National Framework on the Transition to an Open Research Environment*. <https://enterprise.gov.ie/en/Publications/Publication-files/National-Framework-on-the-Transition-to-an-Open-Research-Environment.pdf>
- Grove, J. (2022, August 3). Academic reputation ‘still driven by journal prestige’ – survey. *Times Higher Education (THE)*. <https://www.timeshighereducation.com/news/academic-reputation-still-driven-journal-prestige-survey>
- Hatch, A., & Curry, S. (2020). Changing how we evaluate research is difficult, but not impossible. *eLife*, 9, e58654. <https://doi.org/10.7554/eLife.58654>
- Henriksen, D. (2016). The rise in co-authorship in the social sciences (1980–2013). *Scientometrics*, 107(2), 455–476. <https://doi.org/10.1007/s11192-016-1849-x>
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429–431. <https://doi.org/10.1038/520429a>
- Hsiao, T.-K., & Schneider, J. (2021). Continued use of retracted papers: Temporal trends in citations and (lack of) awareness of retractions shown in citation contexts in biomedicine. *Quantitative Science Studies*, 2(4), 1144–1169. https://doi.org/10.1162/qss_a_00155
- Johns, A. (2000). Miscellaneous methods: Authors, societies and journals in early modern England. *The British Journal for the History of Science*, 33(2), 159–186. <https://doi.org/10.1017/S0007087499003933>
- Kehm, B. M. (2020). Global University Rankings: Impacts and Applications. In M. Biagioli & A. Lippman (Eds.), *Gaming the Metrics: Misconduct and Manipulation in Academic Research*. The MIT Press. <https://doi.org/10.7551/mitpress/11087.003.0009>
- Larivière, V., Haustein, S., & Mongeon, P. (2015). The Oligopoly of Academic Publishers in the Digital Era. *PLOS ONE*, 10(6), e0127502. <https://doi.org/10.1371/journal.pone.0127502>

- Larivière, V., Kiermer, V., MacCallum, C. J., McNutt, M., Patterson, M., Pulverer, B., Swaminathan, S., Taylor, S., & Curry, S. (2016). *A simple proposal for the publication of journal citation distributions* [Preprint]. <https://doi.org/10.1101/062109>
- Lockwood, M. (2020). Editorial: Citation malpractice. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 476(2242), 20200746. <https://doi.org/10.1098/rspa.2020.0746>
- Mallapaty, S. (2020). China bans cash rewards for publishing papers. *Nature*, 579(7797), 18–18. <https://doi.org/10.1038/d41586-020-00574-8>
- Matthews, D. (2019, February 22). *Elsevier profits near £1 billion despite European disputes* / *Times Higher Education (THE)*. <https://www.timeshighereducation.com/news/elsevier-profits-near-ps1-billion-despite-european-disputes>
- McKiernan, E. C., Schimanski, L. A., Muñoz Nieves, C., Matthias, L., Niles, M. T., & Alperin, J. P. (2019). Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations. *ELife*, 8, e47338. <https://doi.org/10.7554/eLife.47338>
- Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation. (2021). *Deuxième plan national pour la science ouverte: Généraliser la science ouverte en France 2021-2024*. https://www.ouvrirelascience.fr/wp-content/uploads/2021/06/Deuxieme-Plan-National-Science-Ouverte_2021-2024.pdf
- Miyakawa, T. (2020). No raw data, no science: Another possible source of the reproducibility crisis. *Molecular Brain*, 13(1), 24. <https://doi.org/10.1186/s13041-020-0552-2>
- O'Carroll, C., Rentier, B., Cabello-Valdes, C., Esposito, F., Kaunismaa, E., Maas, K., Metcalfe, J., Mcallister, D., & Vandeveld, K. (2017). *Evaluation of research careers fully acknowledging Open Science practices: Rewards, incentives and/or recognition for researchers practicing Open Science*. Publications Office. <https://data.europa.eu/doi/10.2777/75255>
- Oransky, I. (2020). Retraction Watch: What We've Learned and How Metrics Play a Role. In M. Biagioli & A. Lippman (Eds.), *Gaming the Metrics: Misconduct and Manipulation in Academic Research* (pp. 141–148). The MIT Press. <https://doi.org/10.7551/mitpress/11087.003.0014>
- Oransky, I. (2022). Retractions are increasing, but not enough. *Nature*, 608(7921), 9–9. <https://doi.org/10.1038/d41586-022-02071-6>
- Piller, C. (2022). Potential fabrication in research images threatens key theory of Alzheimer's disease. *Science*, 377(6604), 358–363. <https://www.science.org/content/article/potential-fabrication-research-images-threatens-key-theory-alzheimers-disease>

- Redden, E. (2019, January 23). *Stepping Out of the Rat Race*. Inside Higher Ed. <https://www.insidehighered.com/news/2019/01/23/ghent-university-belgium-embraces-new-approach-faculty-evaluation-less-focused>
- Rodrigues, E. (2020). A pandemia e a emergência da ciência aberta. In M. Martins & E. Rodrigues (Eds.), *A Universidade do Minho em tempos de pandemia: Tomo II: Re(Ações)* (Vol. 2, pp. 263–294). UMinho Editora. <https://doi.org/10.21814/uminho.ed.24.12>
- Roskams-Edris, D., Poline, JB., Stikov, N. (2019, Agosto 15). The Canadian Open Neuroscience Platform: Catching Up to Plan S and Going Further. *PLOS Blogs*. <https://theplosblog.plos.org/2019/08/the-canadian-open-neuroscience-platform-catching-up-to-plan-s-and-going-further/>, consultado em 3 de junho de 2022.
- Saenen, B., Hatch, A., Curry, S., Proudman, V., & Lakoduk, A. (2021). *Reimagining Academic Career Assessment: Stories of innovation and change*. DORA, EUA, SPARC Europe. https://eua.eu/downloads/publications/eua-dora-sparc_case%20study%20report.pdf
- Saenen, B., Morais, R., Gaillard, V., & Borrell-Damián, L. (2019). *Research Assessment in the Transition to Open Science*. European University Association. <https://eua.eu/downloads/publications/research%20assessment%20in%20the%20transition%20to%20open%20science.pdf>
- Saenen, D. B., & Borrell-Damián, D. L. (2019). *Reflections on University Research Assessment*. European University Association. <https://eua.eu/resources/publications/825:reflections-on-university-research-assessment-key-concepts,-issues-and-actors.html>
- Singh Chawla, D. (2019). Elsevier investigates hundreds of peer reviewers for manipulating citations. *Nature*, 573(7773), 174–174. <https://doi.org/10.1038/d41586-019-02639-9>
- Steen, R. G., Casadevall, A., & Fang, F. C. (2013). Why Has the Number of Scientific Retractions Increased? *PLOS ONE*, 8(7), e68397. <https://doi.org/10.1371/journal.pone.0068397>
- STM. (2021). *STM Global Brief 2021 – Economics and Market Size*. STM. https://www.stm-assoc.org/2021_10_19_STM_Global_Brief_2021_Economics_and_Market_Size.pdf
- Tonta, Y. (2014). *Use and Misuse of Bibliometric Measures for Assessment of Academic Performance, Tenure and Publication Support*. <https://doi.org/10.13140/2.1.3903.3289>
- UNESCO. (2021). *UNESCO Recommendation on Open Science*. <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>
- Universities Norway. (2021). *NOR-CAM – A toolbox for recognition and rewards in academic careers*. https://www.uhr.no/en/_f/p3/i86e9ec84-3b3d-48ce-8167-bbae0f507ce8/nor-cam-a-tool-box-for-assessment-and-rewards.pdf
- Van Noorden, R. (2012). Researchers feel pressure to cite superfluous papers. *Nature*. <https://doi.org/10.1038/nature.2012.9968>

- Van Noorden, R. (2020a). Highly cited researcher banned from journal board for citation abuse. *Nature*, 578(7794), 200–201. <https://doi.org/10.1038/d41586-020-00335-7>
- Van Noorden, R. (2020b). Signs of ‘citation hacking’ flagged in scientific papers. *Nature*, 584(7822), 508–508. <https://doi.org/10.1038/d41586-020-02378-2>
- Van Noorden, R., & Singh Chawla, D. (2019). Hundreds of extreme self-citing scientists revealed in new database. *Nature*, 572(7771), 578–579. <https://doi.org/10.1038/d41586-019-02479-7>
- VSNU, KNAW & NOW (2019). *Room for everyone’s talent—Towards a new balance in the recognition and rewards for academics*.
<https://www.vsnul.nl/files/documenten/Domeinen/Onderzoek/Position%20paper%20Room%20for%20everyone%E2%80%99s%20talent.pdf>
- VSNU, KNAW and NWO. (2020). *The Strategy Evaluation Protocol 2021-2027*.
https://www.universiteitenvannederland.nl/files/documenten/Domeinen/Onderzoek/SEP_2021-2027.pdf
- Wellcome Trust. (2020). *What researchers think about the culture they work in*. Wellcome.
<https://wellcome.org/reports/what-researchers-think-about-research-culture>
- Why high-profile journals have more retractions. (2014). *Nature*.
<https://doi.org/10.1038/nature.2014.15951>
- Wilhite, A. W., & Fong, E. A. (2012). Coercive Citation in Academic Publishing. *Science*, 335(6068), 542–543. <https://doi.org/10.1126/science.1212540>
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J., & Johnson, B. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. <https://doi.org/10.13140/RG.2.1.4929.1363>
- Zhang, L., & Sivertsen, G. (2020). The New Research Assessment Reform in China and Its Implementation. *Scholarly Assessment Reports*, 2(1), 3. <https://doi.org/10.29024/sar.15>