

Texto para subsidiar as colocações das Áreas de Avaliação da CAPES durante a audiência pública virtual convocada pelo MPF-RJ.

### **O Modelo de Avaliação CAPES**

O modelo de avaliação dos Programas de Pós-Graduação vem sendo desenvolvido, aperfeiçoado e aplicado há décadas. Como observou o conselheiro da Sociedade Brasileira para o Progresso da Ciência e membro da Academia Brasileira de Ciências, Professor Carlos Alexandre Netto, ao comentar as propostas de melhoria desse modelo encaminhadas pela ABC e pela SBPC (Alexandre Netto, 2018), toda a comunidade reconhece que foi a sistemática de avaliação periódica dos programas de pós-graduação pela CAPES que levou ao crescimento da produção científica brasileira. Destacou ainda que o grande envolvimento da comunidade acadêmica da pós-graduação com a avaliação, (quase duas mil pessoas na avaliação do quadriênio 2013 a 2016) é um indicador da capilaridade do processo. Este modelo tem por características ser comparativo, formativo, sistematizado, participativo e transparente.

Essa avaliação dos Programas pela CAPES é baseada no modelo referenciado pela comparação e não no modelo referenciado pelos critérios preestabelecidos, conforme a distinção de Popham & Husek (1969). Esse modelo é universalmente utilizado em avaliações de larga escala que resultam no escalonamento dos resultados. É uma forma de assegurar que os diversos níveis de uma escala avaliativa sejam ocupados. Nesse modelo, os indicadores são pré-determinados e divulgados antes do período da avaliação, os critérios, por sua vez, são desenvolvidos comparativamente durante a própria avaliação, a partir dos dados analisados. Já num modelo de definição a priori das notas de corte ou critérios, sem conhecimento do padrão de distribuição dos resultados, Programas com desempenhos diferentes poderiam ser enquadrados dentro das mesmas faixas, resultando em notas iguais.

Vale destacar que a avaliação comparativa representa uma abordagem legítima, amplamente utilizada nacional e internacionalmente, especialmente na medida em que os resultados são apresentados por meio de uma escala vertical que busca diferenciar (ou discriminar) entre diferentes níveis de desempenho. É por isso que todos os *rankings* encontrados na literatura, inclusive os *rankings* de instituições superiores, são baseados em avaliações comparativas e não de critérios predeterminados, pois a lógica de qualquer ranking é sempre diferenciar. A avaliação comparativa é, por exemplo, também utilizada pelo INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira) ao calcular seus indicadores de qualidade de instituições e de cursos de educação superior (as notas do ENADE, do CPC e do IGC, por exemplo). Um outro exemplo de avaliação comparativa é o vestibular para ingresso nas universidades públicas. A partir de 1970, a prova passou a ser classificatória, ou seja, comparativa:

A lógica da avaliação comparativa da CAPES, na qual os resultados são escalonados, se fundamenta no fato de que a escala exerce um papel central no sentido de orientar políticas e decisões no âmbito da pós-graduação brasileira. Criada em 1980 com cinco graus (conceitos) e ampliada em 1998 para sete graus (notas - escala de 1 a 7, onde 1 e 2 representam Programas

não recomendados e 6 e 7, níveis de excelência), a referida escala serve a distintas finalidades. Em primeiro lugar é utilizada pela CAPES para fins administrativos, referenciando decisões a respeito da distribuição de recursos (de diversos tipos) entre os componentes do sistema. A escala facilita a eficiência dessa distribuição, pois ao saber sobre a qualidade de cada programa, a alocação pode ser feita de forma a maximizar sua efetividade seja destacando os Programas que ocupam o topo da escala, seja permitindo o uso de recursos para corrigir assimetrias apoiando aqueles que estão na base da escala. A escala também serve para incentivar a melhoria dos programas de pós-graduação, pois há benefícios de diferentes tipos, tanto materiais quanto simbólicos. O fato de que esse incentivo funciona é demonstrado pela produção científica produzida pelo Brasil, que no decorrer do tempo, se tornou uma das mais altas entre os países do mundo. Uma terceira finalidade da escala adotada pela CAPES é a promoção da internacionalização da pós-graduação. Reservando o topo da escala para Programas de "excelência internacional" sinaliza para professores, pesquisadores e alunos de outros países a qualidade reconhecidamente alta de certos Programas, facilitando colaborações, parcerias e intercâmbios com elementos em todas as partes do mundo que promovem o desenvolvimento científico e tecnológico no sentido global. A escala funciona também para fundamentar decisões de alunos ao determinar seu programa de estudo, de instituições ao selecionar os seus quadros acadêmicos e de muitas empresas ao contratar e promover seu pessoal. Por essas razões, a preservação da integridade da escala de avaliação da CAPES é essencial, e essa preservação exige a utilização do modelo de avaliação comparativo, evitando, dessa forma, a concentração de programas em um único grau ou nível.

A avaliação dos programas é também de natureza formativa, constituindo um processo contínuo de aperfeiçoamento. Ao finalizar um processo de avaliação de permanência, as diferentes Comissões que são estabelecidas junto com os Coordenadores de Área e os Coordenadores de Programa possuem ampla condição de verificar, em função dos avanços obtidos pelos Programas ao longo do período de avaliação, a necessidade de aprimoramento da Ficha de Avaliação e dos indicadores de cada Área. Ressalta-se aqui que, em diversos momentos ao longo das últimas décadas, a comunidade acadêmico-científica, através da Comissão Especial de Acompanhamento do Programa Nacional de Pós-Graduação, tem recomendado à CAPES a necessidade de ajustes/aperfeiçoamentos maiores, como aconteceu no relatório 2018 desta Comissão que surge após consulta a mais de 30 setores e entidades relacionados ao Sistema Nacional de Pós-Graduação como ABC (Academia Brasileira de Ciência), ANDIFES (Associação Nacional dos Dirigentes das Instituições Federais de Ensino Superior), ABRUEM (Associação Brasileira dos Reitores das Universidades Estaduais e Municipais), ABRUC (Associação Brasileira das Universidades Comunitárias), CNE (Conselho Nacional de Educação), CONFAP (Conselho Nacional das Fundações Estaduais de Amparo à Pesquisa), CONSECTI (Conselho Nacional de Secretários Estaduais para Assuntos de CT&I), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CTC-ES (Conselho Técnico-Científico da Educação Superior), FINEP (Financiadora de Estudos e Projetos), FOPROP (Fórum Nacional de Pró-reitores de Pesquisa e Pós-graduação), entre outros, caracterizando o diálogo entre os órgãos públicos, as instituições de ensino superior, as organizações da sociedade civil envolvidas no processo e a comunidade, para consecução de objetivos comuns. Este relatório, aprovado no Conselho Superior da CAPES em 10 de outubro

de 2018, conselho que conta com ampla representação da comunidade acadêmico-científica, incluindo representação estudantil, induziu, entre outras coisas, a reestruturação de quesitos e itens da Ficha de Avaliação, necessários para dar maior clareza à avaliação de resultados e processos. A implementação de aperfeiçoamentos na Ficha de Avaliação com seus indicadores, através deste longo processo cooperativo entre pares e consulta ampla à comunidade, implica obviamente na análise e ajuste dos indicadores das Áreas a favor de uma melhor avaliação dos Programas envolvidos. Não é, portanto, casual o apoio manifesto, e público, destas entidades da sociedade civil e das Instituições de Ensino Superior, como ABC, FOPROP e SBPC, entre outras, à manutenção e aperfeiçoamento do sistema de avaliação da CAPES.

É importante reconhecer que o processo de preparação de uma avaliação quadrienal ou de permanência (e também das avaliações de propostas de novos cursos de pós-graduação stricto sensu - APCN), se inicia comumente bem antes da mesma. Os procedimentos se iniciam com Comissões que avaliam resultados de período/ciclo avaliativo anterior, há ampla consulta à comunidade acadêmico-científica, os resultados e sugestões do trabalho destas Comissões são repassados à CAPES que então inicia os procedimentos de aperfeiçoamento visando atender aos anseios da sociedade através da implementação de Grupos de Trabalho. Estes fazem recomendações que são assimiladas pelas Áreas de Avaliação e Programas de Pós-Graduação que em reuniões conjuntas com os coordenadores de Programas nos Seminários de Meio Termo (para os quais são convocados os Coordenadores de todos os Programas de Pós-Graduação das 49 Áreas de Avaliação), discutem os aperfeiçoamentos realizados nos indicadores e os validam utilizando dados dos dois primeiros anos do ciclo avaliativo. Só após esta ampla consulta e debate são concluídas e publicadas as Fichas de Avaliação das 49 Áreas e posteriormente ainda o regulamento que operacionaliza os procedimentos da avaliação quadrienal a ocorrer no Período de Avaliação. Nessa direção, o sistema de avaliação de programas pela CAPES caracteriza-se também como participativo, sistematizado e transparente.

Quanto à transparência, vale ressaltar que as atuais Fichas de Avaliação, que descrevem os indicadores aperfeiçoados para a avaliação de permanência (ciclo avaliativo 2017-2020), para cada item e Quesito, foram publicadas e amplamente divulgadas entre final de 2019 e início do segundo semestre de 2020. O Período de Avaliação de acordo com calendário iniciou-se após a Coleta de dados dos Programas que, para o ano base 2020, se encerrou em 14 de junho de 2021. Este intervalo, dentro do qual houve ainda a oportunidade dos Programas efetuarem ajustes e mudanças no Coleta de anos anteriores (ReColeta), de quase um ano seria momento adequado, caso houvesse necessidade, de interpor à CAPES algum tipo de questionamento sobre a inadequação de algum quesito/item/indicador. Observamos que a CAPES não recebeu neste intervalo, entre início de 2020 até segundo semestre de 2021, nem nos intervalos similares de ciclos avaliativos anteriores, qualquer questionamento quanto aos aperfeiçoamentos de Ficha de Avaliação e seus indicadores. Trata-se aqui, claramente, de entendimento amplo da parte dos Programas de Pós-Graduação e de suas Pró-reitorias que os ajustes visam uma melhor qualidade no processo de avaliação o que os favorece. O aperfeiçoamento dos instrumentos Ficha de Avaliação e Regulamento da Avaliação é parte de uma cultura acadêmico-científica de avaliação por pares entendida como benéfica para uma melhor avaliação formativa e

indutora de desenvolvimento dos Programas. A destacar ainda que aperfeiçoamentos mais importantes de indicadores (ex. indicadores novos), além de passarem pelo processo de discussão descrito acima, são introduzidos com peso reduzido visando induzir os aperfeiçoamentos necessários gradualmente sem alterar os resultados da avaliação. Novamente é esta característica participativa, inclusiva, com características fortes de autoavaliação, que leva os segmentos envolvidos, Programas de Pós-Graduação e suas associações como a ANPOCS (Associação Nacional de Pós-Graduação e Pesquisa em Ciências Sociais), a ANPED (Associação Nacional de Pós-Graduação e Pesquisa em Educação), a ANPET (Associação Nacional de Pesquisa e Ensino em Transportes), associações e fóruns de coordenadores de programas de pós-graduação, para citar apenas alguns, e mesmo o Fórum de Pró-reitores de Pós-Graduação, FOPROP, a se manifestar publicamente em defesa do sistema de avaliação construído e aperfeiçoado ao longo dos últimos mais de 40 anos.

Na verdade, o aperfeiçoamento constante da avaliação da CAPES vem garantindo nos últimos quadriênios forte estabilidade de notas, com baixíssimo número de Programas que sofrem rebaixamento de nota ou descredenciamento (10% ou menos) quando comparados aos que mantêm ou sobem de nota (90% ou mais). A imprevisibilidade surgiria agora caso haja a necessidade de utilizar os procedimentos e instrumentos aplicados na avaliação de permanência realizada em 2017 que se encontram claramente defasados e descolados dos dados de realidade coletados pelos Programas de Pós-Graduação e dos instrumentos acessórios de avaliação, como a de qualificação de periódicos.

Nesse sentido, é importante enfatizar os prejuízos que serão sofridos pelos Programas de Pós-Graduação caso os indicadores e critérios do período 2013-2016 sejam aplicados na avaliação ora em andamento. Isso porque sendo a avaliação da CAPES um processo constante e contínuo, seguindo uma mesma dinâmica desde sua implementação em 1980, os Programas sabem que o que vale é a ficha construída coletivamente e divulgada a partir dos Seminários de Meio Termo, realizados durante o quadriênio. Os Programas preparam seus relatórios com base na ficha mais atual, uma vez que esse tem sido o procedimento adotado historicamente. A retomada de indicadores do passado, muitas vezes aperfeiçoados (ou até eliminados) no quadriênio em vigor, aproveitando reflexões sobre a avaliação anterior, geraria um processo avaliativo não esperado e, assim, prejudicial em relação aos programas envolvidos.

Vale destacar ainda que o funcionamento dos Programas é contínuo (projetos iniciados ao longo de um quadriênio resultam comumente em orientações e produções intelectuais que podem ser concluídas apenas no período de avaliações subsequentes), quando foram examinados os resultados da avaliação de permanência de 2017 (referente aos anos de produção 2013 a 2016) que permitiram ou indicaram a necessidade de aperfeiçoamentos nos indicadores para atender às mudanças nos padrões de produção de conhecimento e formação de Mestres e Doutores já haviam decorrido praticamente dois anos do atual período de avaliação (2017-2020). Utilizar estes resultados apenas para a avaliação subsequente à atual, potencialmente apenas para a avaliação de permanência de 2025, — ou seja, já completamente defasados — colocaria um atraso de quase oito anos para estimular, através destes

aperfeiçoamentos, as melhores práticas nos processos de formação e pesquisa, o que certamente contraria o interesse público dada sua ineficiência.

De forma interessante e falha em sua tréplica, o MPF-RJ associa o sistema de avaliação a um jogo. Textualmente afirma: “Não se nega que possa haver aperfeiçoamentos, avanços, na avaliação. O problema é quando se mudam as regras do jogo depois de começado, privilegiando uns em detrimento de outros programas de pós-graduação.”

Na realidade, vários estudos (Butler, 2002, 2003a, 2003b, 2004; Moed, 2008; Pires, 2020) indicam que há fortes riscos em estruturar o sistema de avaliação aproximando-o a um “jogo”. O “Leiden Manifesto”, publicado na prestigiosa revista Nature (Hicks et al., 2015), alerta para os potenciais efeitos negativos do uso de indicadores na avaliação da pesquisa científica que podem convidar a jogos e desvios de metas (onde o que será medido torna-se o objetivo). Neste sentido, Hicks e colaboradores (2015) sugerem que os indicadores devem ser regularmente avaliados e atualizados. Os objetivos da pesquisa e aqueles da avaliação mudam e o sistema de pesquisa, ele mesmo, evolui. Assim, concluem, métricas anteriormente adequadas se tornam obsoletas e novas surgem. Sugerindo ainda que os sistemas de indicadores devam ser continuamente revisados e se necessário modificados.

De acordo com Rijcke e colaboradores (2016), desde o desenvolvimento e primeiras aplicações de indicadores de performance na pesquisa nos anos 70 surgiram as discussões sobre o comportamento estratégico e “jogar para o sistema”. Sistemas de avaliação influenciam o comportamento dos pesquisadores de duas formas. A primeira acima citada é deslocar os objetivos - obter boa avaliação nos critérios passa a ser o objetivo. A segunda é uma transformação ainda mais preocupante pela qual os critérios de avaliação afetam o processo de fazer ciência por si mesmo (por exemplo, evitando risco na seleção de tópicos de pesquisa). Espland e Sauder em 2007 conceituaram esses mecanismos como de “reatividade”, em que a ideia básica é que os indivíduos alteram seu comportamento em reação a serem avaliados, observados ou medidos.

Um sistema no qual os critérios de avaliação não acompanhem a evolução e avanços do sistema de produção científica e de formação de mestres e doutores poderá, na forma de um “jogo”, deslocar objetivos para alcançar as metas propostas nos indicadores, desvirtuando a produção de conhecimento científico e impactando seu potencial inovador. Quanto à alegação do MPF-RJ de “privilegiar uns em detrimento de outros programas de pós graduação” deve ser alertado que a divulgação antecipada e não aperfeiçoada de indicadores e critérios certamente desencadeará um aumento na assimetria regional e intrarregional ou entre notas do sistema no que se refere à qualificação dos Programas uma vez que, aqueles mais estruturados e com maior disponibilidade de recursos humanos e acesso a financiamento (normalmente os que já possuem melhores notas) saberão/poderão usar mais rapidamente as regras/indicadores e alcançar com maior facilidade os “novos objetivos”. Marques e colaboradores (2017) a este respeito indicam que é esperado que as elites científicas, com acesso a mais recursos, e reputações estabelecidas dentro das hierarquias acadêmicas, aprendam rapidamente as regras e normas de um sistema de avaliação e depois as manipulem estratégica e taticamente para

maximizar suas vantagens. Tal fato reforça o “efeito Mateus” na ciência (Merton, 1968), aumentando a distância dos que têm mais recursos para aqueles menos providos. Assim, o sistema de Pós-Graduação não é e não pode ser reduzido a um “jogo” e a sua avaliação não deve, portanto, se basear num sistema de “regras anteriores ao início do jogo”. Na realidade, dificilmente um coordenador ou docente de Programa de Pós-Graduação desconhece o que deve ser feito para melhorar a nota de seu Programa: formar Mestres e Doutores com qualidade e publicar os resultados de suas pesquisas para avanço da ciência e da tecnologia.

Por fim, os fatos, os relatos, os resultados e, em especial, os estudos científicos, comprovam a oportunidade e a adequabilidade deste modelo comparativo, formativo, sistematizado, participativo e transparente de Avaliação dos Programas de Pós-Graduação utilizado pela CAPES há mais de 40 anos. Um modelo estável, como é desejável, mas não inerte. Que dialeticamente interage, confronta marcos e potenciais avanços, e se transforma. Evolui, como tudo na ciência e na vida. Somente assim, nesta contínua metamorfose, um modelo de avaliação é capaz de apoiar a inovação e o desenvolvimento da Pós-graduação no País.

## **Bibliografia**

ALEXANDRE NETTO, C. Princípios para um novo modelo de avaliação da pós-graduação. *Ciência & Cultura*, S. Paulo, v. 70 n.3 p. 47-51, 2018.

BUTLER, L. A list of published papers is no measure of value. *Nature*, London, v. 419, n. 6910, p. 877, 2002.

BUTLER, L. Explaining Australia’s increased share of ISI publications—the effects of a funding formula based on publication counts. *Research Policy*, Amsterdam, v. 32, n. 1, p. 143-155, 2003a.

BUTLER, L. Modifying publication practices in response to funding formulas. *Research Evaluation*, Oxford, v. 12, n. 1, p. 39-46, 2003b.

BUTLER, L. ‘What Happens When Funding is Linked to Publication Counts?’. In: MOED, H. F.; GLANZEL, W; SCHMOCH, U (eds.). *Handbook of Quantitative Science*. 2004.

ESPELAND, W.; SAUDER, M. Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, Chicago, v. 113, n. 1, p. 1-40, 2007.

HICKS, D. et al. Bibliometrics: the Leiden Manifesto for research metrics, *Nature*, London, v. 520, n. 7548, p. 429-431, 2015.

MARQUES, M. et al. How does research evaluation impact educational research? Exploring intended and unintended consequences of research assessment in the United Kingdom, 1986–2014. *European Educational Research Journal*, Thousand Oaks, v. 16, n. 6, p. 820-842, 2017.

MERTON, R.K. The Matthew effect in science. The reward and communication systems of science are considered. *Science*, v. 159, n. 3810, p. 56-63, 1968.

MOED, H.F. UK Research Assessment Exercises: Informed judgments on research quality or quantity? *Scientometrics*, Berlin, v. 74, n. 1, p. 153-161, 2008.

PIRES, A.S. A framework of the monitoring scientific production behavior in research evaluation systems based on journal ranking lists: the case of Brazil. Tese de Doutorado em Educação, Universidade Federal de Rio Grande do Sul, 2020.

POPHAM, W.J.; HUSEK, T.R. Implications of Criterion-Referenced Measurement. *Journal of Educational Measurement*, v. 6, n. 1, pp. 1-9, 1969.

RIJCKE, S. et al. Evaluation practices and effects of indicator use – a literature review. *Research Evaluation*, Oxford, v. 25, n. 2, p. 161-169, 2016.