



**AVALIAÇÃO DA PRODUÇÃO DOCENTE NO ENSINO SUPERIOR:
POSSIBILIDADES E LIMITES**

CHIZZOTTI, Antonio

Docente do Programa de Pós-Graduação
em Educação: Currículo - PUC/SP.
Departamento de Fundamentos da Educação.
Email. anchizo@uol.com.br.

PONCE, Branca Jurema

Docente do Programa de Pós-Graduação
em Educação: Currículo - PUC/SP.
Departamento de Fundamentos da Educação.
Email. tresponces@uol.com.br



RESUMO

O artigo analisa o debate sobre a avaliação da produção docente no ensino Superior focando, de modo especial, a avaliação nos Programas de Pós-Graduação no Brasil. Para isso, dois paradigmas de avaliação são abordados: avaliação pelos pares e avaliação pela quantificação das citações nas bases de dados e das classificações dos veículos de publicação (editoras, periódicos, apostilas etc.). Conclui-se que a avaliação por quantificação suplantou a tradição avaliativa qualitativa dos docentes do ensino superior. Cada vez mais, tem-se utilizado parâmetros numéricos para dimensionar a produção e a qualidade dos docentes. A quantificação encontra suporte nas bases de dados disponíveis e em sistemas de classificação dos meios que publicam as produções acadêmicas. Os processos avaliativos provocam debates tanto em relação à pertinência quanto à validade dos seus critérios. Nesse debate, expressa-se a posição dos que defendem um padrão de gestão do ensino superior apoiado em resultados quantificáveis e, em contrapartida, é veemente a oposição de pesquisadores a esses critérios, pela falibilidade e impertinência desses para as ciências humanas e sociais.

Palavras-chave: ensino superior – paradigmas de avaliação - avaliação pelos pares – avaliação bibliométrica – avaliação de produção científica

ABSTRACT

The article discusses the assessment approaches used to analyze university lecturers academic and scientific production, particularly from those who work in post-grad courses in Brazil. For that, two assessment paradigms are used: peer review and citations frequency in scientific data base, as well as classification procedure by quality of means of publication. Results reveal that quantitative assessment outnumber the traditional qualitative approach. It seems that quantitative analysis has been more frequently used nowadays. Such analysis has become more feasible due to easy access to data base and classification systems. The assessment processes trigger debates about how valid and adequate the chosen set of criteria used are. The debates involve those who are for and against quantitative procedures, and the latter argues that such criteria may not be suitable for humanities and social sciences.

Key words: higher education – assessment paradigms - peer review – bibliometric evaluation - scientific production assessment

1. INTRODUÇÃO

A avaliação da produção acadêmica docente em ciências humanas e sociais está no centro de um amplo debate, nacional e internacional, sobre os processos de avaliação da educação superior e seus critérios, e sobre as concepções de avaliação da produção e da qualidade das atividades científicas dos professores. Observa-se que o tema está presente em todas as reuniões das associações regionais e nacionais de docentes quando se analisam as recomendações ou os resultados da avaliação do Comitê de Acompanhamento da Avaliação da Área. O resultado final e a decorrente atribuição de uma classificação têm implicações relevantes na qualificação institucional dos Programas de Pós-Graduação, na distribuição de apoios, bolsas e benefícios acadêmicos. Essas consequências mobilizam posições calorosas (e mais candentes se mostram), quando são tomadas como qualificações autorizadas da produtividade e qualidade docente.

A avaliação fundamenta-se em um juízo de valor e sempre evoca as concepções teóricas sobre a avaliação, os pressupostos em que se fundamentam, os critérios adotados, as finalidades perseguidas, os resultados alcançados e os usos da avaliação. É, necessariamente, um campo envolto em questões teóricas controversas; mas, o juízo final tem um cunho aprobatório, ou não, do percurso de vida do docente; mobiliza, por isso, emoções fortes, tensões e discursos ardorosos. Essa convulsão avaliativa, não é uma particularidade doméstica; está presente nas avaliações nacionais, reproduzindo questões que ocorrem no âmbito internacional, sobre um aspecto da docência no ensino superior: o que faz o docente e o que se espera dele ou quais publicações levar em conta para avaliar a produtividade e a qualidade científica? (REY, 2010).

A avaliação docente, antes limitada aos processos de admissão e promoção na carreira acadêmica, ganhou foros polêmicos nos anos 1980, a partir do movimento de avaliação das universidades e da criação de organismos nacionais de avaliação institucional do ensino superior.

Sempre houve uma avaliação informal das universidades, calcada no prestígio de seu corpo docente, nos prêmios científicos da instituição, na vetustez da sua fundação, no volume de recursos, no número de alunos, na quantidade e qualidade arquitetônica dos edifícios, etc. O processo de globalização, com o reconhecimento da abertura de um amplo mercado, aberto pela demanda de novos públicos, necessidades emergentes, internacionalização de cursos acadêmicos e o estímulo à competição das instituições universitárias, provocou o fortalecimento de dispositivos sistemáticos de avaliação institucional e a emergência de

padrões comparativos competitivos, dentre os quais a produtividade e a qualidade do corpo docente. A avaliação docente, ou melhor, um determinado tipo de avaliação – decisivamente competitivo - ingressou na pauta das avaliações institucionais

Em 1985, a França cria um Comitê Nacional de Avaliação das Universidades, sugerido pelo relatório solicitado pelo governo sobre o ensino e o desenvolvimento científico. Laurent Schwartz, o presidente do comitê, havia sido o responsável pela publicação de *Para salvar a universidade* (1983), em que faz um diagnóstico do ensino e propõe sugestões para a avaliação das instituições universitárias. A avaliação institucional ganhou cada vez mais destaque nas políticas educacionais e constituiu um dos objetivos do Ministério da Educação do Brasil, especialmente no final dos anos 1980 e na década de 1990. Nos anos seguintes, estenderam-se as classificações e avaliações às produções dos docentes.

As controvérsias em torno da avaliação acadêmica estão polarizadas entre **duas concepções básicas** de avaliação da produção acadêmica docente: uma, historicamente utilizada na vida universitária, apóia-se na avaliação pelos pares (*peer review*); outra, ascendente, recente e cada vez mais dominante, recorre aos recursos mensuradores para a quantificação da produtividade para, por meio dela, inferir uma concepção determinada de qualidade nas atividades científicas docentes. As duas concepções mencionadas pressupõem critérios avaliativos consequentes a elas, dos quais defluem exuberantes controvérsias polêmicas sobre os objetivos e usos da avaliação. Apresentam-se, ora contrapondo-se com discursos veementes, ora invocando a complementaridade mútua, a fim de, nesse caso, com dupla referência judicial, alcançar uma avaliação comensurável consistente e convincente.

A avaliação supõe um juízo de valor sobre a produção acadêmica do docente, apoiando-se em diferentes indicações que demonstrem a dedicação ao ensino, o empenho em pesquisa e em outras atividades de gestão, de produção e difusão científicas. Pressupõe que, confrontando os recursos necessários, os objetivos assumidos, as ações realizadas e os resultados finais alcançados, possa-se apreciar a eficácia do corpo docente. Quando essa avaliação está inserida em uma política pública de avaliação, o julgamento de valor pressupõe um processo cognitivo rigoroso, um processo formal, sujeito a normas definidas pelas instâncias avaliadoras, e impõe a mobilização de um aparato coerente de procedimentos para garantir a qualidade.

2. REFERENCIAL TEÓRICO

2.1. Produção avaliada pelos pares (peer reviewed)

A tradição acadêmica, apoiada na autonomia da universidade, reconhece que a avaliação pelos pares é um processo adequado e consistente de avaliação da qualidade científica. A avaliação pelos pares consiste na leitura “anônima” do produto científico de um candidato por outro docente do mesmo campo disciplinar que, de posse de alguns critérios, previamente estabelecidos, emite um juízo sobre a qualidade do produto científico. Pressupõe a isenção rigorosa dos membros da comunidade científica, a competência específica na área de conhecimento, o discernimento judicante, a capacidade de reconhecer o estágio atual e as tendências inovadoras de pesquisas em seu campo de conhecimento e a prudência ética para ajuizar sobre pessoas e produtos científicos.

A avaliação pelos pares é constituída, em geral, por mais de um membro. Quando conflitantes as avaliações entre dois, um terceiro membro deve dirimir a pendência. Esse expediente de triangulação assegura a independência e a isenção do juízo, atributos fundamentais da avaliação fidedigna. Essa sistemática, estendida a outras atividades docentes, configura a qualidade da instituição ou da unidade acadêmica. Essa forma avaliativa é bastante antiga e fartamente conhecida, portanto, não é novidade na vida acadêmica; pelo contrário, os docentes estão habituados a processos avaliativos de sua atividade, seja para a progressão acadêmica, seja para admissão de trabalhos em eventos científicos. Sabem que seu trabalho pode ser aceito ou recusado em razão de objetivos específicos ou de uma concepção do avaliador ou de agência financiadora. Os concursos, com todas as normas rigorosamente seguidas, critérios públicos e corpo docente avaliador competente, são processos avaliativos públicos, muitas vezes polêmicos: mobilizam recursos jurídicos sobre a validade de atos, a transparência procedimental do concurso, a legitimidade dos avaliadores e do resultado final do concurso. Os recursos cabem, por direito consuetudinário e autonomia universitária, às instâncias acadêmicas; mas, algumas vezes, transpõem o recinto universitário e acabam, por vezes inadequadamente, acionando os tribunais. Essas situações de conflito, ainda que indesejáveis, fazem parte da vida democrática, se reconhecido o princípio do contraditório de qualquer processo jurídico. Tais circunstâncias têm sido invocadas por outras concepções avaliativas como evidência da falta de primor de uma avaliação objetivamente rigorosa e apelam para avaliações com parâmetros mensuradores mais consistentes.

A avaliação da produção docente nos programas de pós-graduação no Brasil sempre esteve nas pautas de discussões. A Capes desenvolveu, ao longo de anos, sua metodologia. Ela é pioneira na implantação da avaliação da qualidade na pós-graduação *stricto sensu* brasileira e na aplicação dos resultados para reconhecer programas, recomendar melhorias e a eles atribuir nível classificatório. O conceito, os objetivos e as metodologias de avaliação da Capes evoluíram, no curso dos anos, atendendo interesses e critérios de desenvolvimento específicos de áreas de conhecimento, ascendência de grupos hegemônicos e políticas oficiais para o ensino superior. O histórico dessa evolução, revelador das tendências e tensões que a acompanharam na área de educação já foi, em parte, feito. Não cabe evocá-lo, nesse artigo. Mas, a busca de parâmetros para avaliar e gerir o crescimento da pós-graduação *stricto sensu* reforçou a tendência de se adotar parâmetros de quantificação, mais consensuais nas ciências naturais, e por apropriação, estendidos às ciências humanas e sociais.

2.2. Produção avaliada pela quantificação

Essa concepção confia que a avaliação do trabalho de pesquisadores deve apoiar-se em medidas quantitativas por meio de publicações que visam a mensurar a produtividade científica de docentes-pesquisadores. Acredita que a avaliação pelos pares pode ficar confinada ao recinto universitário e exposta ao jogo de influências e interesses de grupos. Sua compreensão é de que a avaliação intramuros acaba reforçando hábitos, vieses e vícios que se estruturam no cotidiano acadêmico e acabam autenticando práticas, grupos e pesquisas que se autovalidam, dos quais não se conhecem os resultados ou, quando se conhecem, não se identifica a relevância científica e, sobretudo, social, dessas atividades. Elas estariam confinadas aos seus autores ou às agências de fomento, sem controle social efetivo ou sem acesso aos resultados alcançados pelos possíveis beneficiários. Para essa concepção, o julgamento pelos pares, quando próximos, está exposto a todas as vicissitudes emocionais do juiz e às influências apaixonadas do juízo; quando remotos, não deixa de atuar com apoios corporativos, acordos tácitos que vigoram no cotidiano. Em suma, a avaliação, nesse caso, perderia vigor, pela falta de suportes objetivos, por respaldar-se nas justificativas subjetivas do avaliador, nublando uma apreciação congruente. Para superar os considerados equívocos, a avaliação tem que encontrar referências em dados objetivos mensuráveis para alcançar uma ponderação consistente.

A tendência crescente e cada vez mais dominante nas classificações comparativas internacionais sobre a avaliação docente revela a busca de parâmetros quantificáveis para avaliar a produtividade acadêmica e estabelecer os indicadores da produtividade e da qualidade dos produtos científicos. Há diversas formas de reunir o fluxo das informações, das comunicações e da difusão do saber nas áreas de conhecimento e sintetizar a evolução de um campo científico. Embora o objetivo seja a condensação de informações sobre a produção de textos e resumos disponíveis, o fluxo de informações, reunidos em bases de dados, tem extrapolado sua finalidade específica e tem sido, cada vez mais, usado para dimensionar a produtividade docente e estabelecer uma competição de caráter quantitativo nos meios acadêmicos. O recurso à bibliometria e suas variantes, o uso das bases de dados, aliados ao regime de classificações dos meios de difusão, têm sido utilizados para aquilatar a produtividade e criar *ranking* de instituições e “recompensas” para os docentes.

Assim, a bibliometria, a quantidade de citações, os prêmios, as condecorações, os trabalhos apresentados em livros, periódicos, editoras e eventos, previamente classificados, apóiam-se na presunção de que as produções mensuradas quantitativamente foram antes também avaliadas qualitativamente. Nesse caso, portanto, admite-se a ilação e as produções são tomadas como qualificadas. Com essas referências, podem-se estabelecer os indicadores para a classificação tanto das instituições universitárias como da produtividade de seu corpo docente.

Para dimensionar a quantidade de produtos, surgiram a bibliometria, a cientometria, a webometria, e estão em franca expansão outros meios mensuradores da produção científica, com o propósito de indexar os artigos e autores de uma seleção de periódicos considerados pertinentes, inscritos em grandes bases de dados, a partir dos quais se pode extrair o número de citações e, por meio dos quais, tem sido avaliada a produtividade docente.

A bibliometria (*bibliometrics*) foi criada em 1969 como a aplicação dos métodos matemáticos e estatísticos para descrever e quantificar o processo de comunicação escrita e fazer prognósticos sobre o desenvolvimento da produção escrita em livros e periódicos. Um dos expedientes utilizados para medir a produção é quantificar o número de citações bibliográficas de um texto científico. As citações são referências bibliográficas que evidenciam os vínculos entre autores de publicações, estabelecem a interrelação de textos e mostram os elos entre pesquisadores, instituições e áreas de conhecimento. A partir do elenco das citações, a bibliometria extrai o volume e o entrelaçamento dos textos citados, infere o desenvolvimento de um campo científico, enuncia o estágio das publicações e a autoria das contribuições dadas em uma área de investigação, respaldando-se na frequência da citação das

palavras, das obras e dos autores. No âmbito da bibliometria desenvolveu-se um campo distinto de mensuração da produção científica e tecnológica, a cientometria (*scientometrics*), para estimar a relevância de periódicos em dada área de conhecimento, pressupondo que a quantidade forma um núcleo de conhecimento científico e, por ilação, de maior relevância e qualidade em um campo do conhecimento. Esse suposto permite estender a quantidade de referências ao maior prestígio e produtividade dos autores: os que mais produzem, presuntivamente têm maior reconhecimento, melhores recompensas acadêmicas, prestígio mais elevado e são guindados ao ápice de uma escala classificatória. A mensuração da quantidade dos produtos constrói os parâmetros de relevância e qualidade da produtividade docente

Com a finalidade de disponibilizar a produção científica, criaram-se e disseminaram-se métodos de levantamentos bibliográficos. Com o avanço da internet, expandiram-se os meios, as formas de triagem e a velocidade na disponibilização das informações, reunidos em bases de dados, originando grandes empresas editoras de indexação da produção científica. Essas bases de dados oferecem diferentes produtos e meios de acesso ao elenco de citações. Diferem muito, ainda, quanto ao modo de disponibilizar os dados e quanto ao custo de uso. Elas também se transformaram em referência para aquilatar a produtividade acadêmica docente. Uma das primeiras iniciativas de reunir as publicações informatizadas disponíveis foi a *Web of Science* (WOS), surgida nos anos 1950.

Webometria (*webmetrics*) não é uma classificação de qualidade acadêmica por meio da quantidade, número e visibilidade das pesquisas; o critério é a disponibilidade da produção científica através da internet. A *Web of Science* tornou-se uma base de dados multidisciplinar que indexa os periódicos e as citações em suas respectivas áreas e alcançou proeminência quase única na área. É um índice de citações, informando, para cada artigo, os documentos por ele citados e os documentos que o citaram. A quantidade de citações é utilizada para medir a produtividade.

Em 1964, o [Institute for Scientific Information](http://www.isinet.org/) (ISI), criou o primeiro índice de citação dos artigos publicados em periódicos acadêmicos, O *Science Citation Index* (SCI) -reunindo citações de 1945 até o presente- e, depois, o *Social Sciences Citation Index* (SSCI)-com citações desde 1956- e o *Arts & Humanities Citation Index* (AHCI)-com citações desde 1975, veicula também o *ISI Web of Knowledge* (WoK), todos propriedade da empresa Thomson Reuters, que comercializa a indexação, mediante subscrição, disponível em impressos, em cd e acessível na *Web of Sciences*.

Outra base de dados de referências bibliográficas em ciências humanas comercializada por empresa é o *Scopus* criado, em 2004, pela empresa holandesa Elsevier. Indexa artigos de cerca de 15.000 periódicos que tenham sido reconhecidos por instituição acadêmica. Apresenta-se como a maior base de resumos e referências bibliográficas de literatura científica revisada por pares.

Em 2004, surge o [Google Scholar](#) (GS), *Google Acadêmico*, auxiliando na identificação da produção relevante do mundo acadêmico. Trata-se de uma base de dados de livre acesso, apresentada como uma fonte de informações gratuitas sobre a literatura acadêmica: artigos revisados por especialistas (peer-reviewed), teses, livros, resumos, artigos integrais, citações de instituições acadêmicas, organizações profissionais, bibliotecas de pré-publicações, universidades e outras entidades acadêmicas.

O portal CAPES, também, disponibiliza, com acesso livre e gratuito, o banco de teses e dissertações brasileiras. O Portal Periódicos CAPES (www.periodicos.capes.gov.br) oferece também acesso livre e gratuito para usuários de instituições cadastradas, aos resumos de 126 bases de dados e aos textos completos de artigos selecionados de 15.475 revistas internacionais, nacionais e estrangeiras. As bases de dados reúnem trabalhos afetos às ciências naturais, humanas e sociais, dentre as quais destacam-se o *Web of Knowledge*, *Scopus* e o *SciELO*.

O *SciELO* (Scientific Eletronic Libray Online), criado pela FAPESP em parceria com o Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (BIREME), abriga, atualmente, 637 periódicos de 8 países ibero-americanos, dos quais 197 são do Brasil, indexados na base de dados da *WEb of Science*. O Portal não pretende ser um instrumento de avaliação da produção docente, mas a inclusão de bibliografia nele é utilizada como referência da excelência da produção e dos autores.

Cada base ou banco de dados oferece diferentes produtos, além de meios próprios de acesso às citações, dispostas de modo diferente, obrigando o usuário comum a um árduo tirocínio para ter acesso às informações ou exigindo um especialista na triagem. Em geral o acesso é feito por meio de um software como o *publish or perish* ou o [Sistema Gerenciador de Banco de Dados](#) (SGBD). Os bancos de dados estão em franca expansão, criando, cada vez mais, bancos específicos; são, também, cada vez mais, a referência para avaliar a produção docente.

Estudos comparativos entre os principais bancos de dados mostram algumas questões: as subscrições têm um alto custo e, em geral, só podem ser bancadas por instituições

universitárias. Algumas são disponibilizadas, mediante senha, para uso em computadores pessoais dos pesquisadores e outras dão acesso somente no ambiente de pesquisa da universidade. Proliferam, felizmente, bancos de dados com acesso livre, atendendo à Iniciativa de Budapeste (*Budapest Open Access Initiative*), que preconiza a livre e irrestrita disponibilização da literatura científica. Entre os bancos de dados com livre acesso, destaca-se o *Google Scholar*, que tende, inclusive, a suplantiar os bancos de dados pagos. Além da gratuidade, conta com espectro maior de triagem de trabalhos científicos, indexando anais de congressos e comunicações em colóquios científicos e ainda, produtos oriundos de canais mais diversificados que o das revistas acadêmicas com comitê editorial.

2.3. As classificações baseadas na quantificação das produções

As bases de dados classificatórias por meio da quantidade de produtos científicos propiciaram o crescimento do *ranking* de instituições, grupos de pesquisa e de pesquisadores. As classificações emergem no bojo da globalização do ensino superior, como um novo pilar de gerência da vida acadêmica, assumidas por instituições universitárias, por governos e pela mídia, como uma referência de excelência da instituição. Essa tendência classificatória surge com a *Classificação de Shangai*, (*Academic Ranking of World University – ARWU*), em 2003. A pedido do governo chinês, a universidade Jiao-Tong elaborou uma classificação em algumas áreas científicas, com a finalidade de enviar alguns alunos excepcionais para as melhores universidades estrangeiras. O autor, um químico responsável pela classificação, baseou-se em dados considerados ‘objetivos’, tais como prêmios Nobel e publicações em revistas internacionais de prestígio. A classificação de Shangai privilegia o resultado de pesquisas e a quantidade de publicações em detrimento do ensino. Com isso, consagra um tipo de universidades, aquelas que teriam se tornado centro de excelência, e subalterniza outros quesitos fundamentais para a vida acadêmica. Essa classificação é difundida, anualmente, pela internet e é reproduzida, pelos meios de comunicação de massa, como uma forma classificatória das melhores universidades, tendo a presunção de ser uma fonte confiável de informações sobre a qualidade da universidade e auxiliar a opção dos “consumidores” de formação superior: estudantes e suas famílias, empregadores públicos ou privados. Os critérios de classificação foram tomados como exemplares para o funcionamento de um novo gerenciamento do ensino superior e, como tal, incorporados pelos departamentos e universidades, pelas empresas e pelo poder público com o objetivo de atribuir

reconhecimento, distribuir recursos ou privilegiar financiamentos. Dispor de imagem internacional permite garantir a atratividade necessária para atender a expansão de um mercado estudantil, nacional e estrangeiro, concentrar talentos, desenvolver convênios e programas científicos interinstitucionais de cooperação internacional, conceder reconhecimento de títulos e justificar o uso de financiamentos públicos ou privados.

As classificações são um ingrediente a mais de uma lógica das relações crescentes entre a educação e o mercado e um novo modo de regulação dos sistemas educativos, controlados pelo discurso de uma nova forma de gestão pública, baseada na racionalidade econômica, eficácia competitiva, gerência da performance, rendimento e retorno dos investimentos.

A classificação acirrou a montante competidora entre universidades, pesquisadores, periódicos, eventos, editoras, compelidos a elevar o respectivo nível classificatório. Os governos favorecem o acirramento competitivo, procurando construir uma boa imagem internacional, seja para sobrelevar a excelência de seu sistema de ensino e ampliar o grau de atratividade diante dos melhores estudantes ou pesquisadores, seja para criar novas oportunidades de cooperação e intercâmbios internacionais científicos e técnicos. Em algumas circunstâncias, transformam-se em obsessão os esforços para multiplicar expedientes que possam elevar o nível classificatório, em detrimento da produção pertinente, paciente, decente e socialmente relevante.

A classificação de Shangai foi questionada pelos critérios adotados, pela valorização exclusiva de algumas áreas, pela visão limitada da universidade e pelos objetivos a que se propôs. Uma classificação que, pelos critérios, releva o estilo universitário dos Estados Unidos e do Reino Unido. Essa classificação provocou, por inconformidade quanto ao seu produto, pelo apoio em outras referências da vida universitária ou pela adoção de outros critérios e metodologias, o surgimento de outras classificações, baseadas em parâmetros mais pertinentes às diversas tradições acadêmicas. O Centro Europeu do Ensino Superior (Center of Higher Education Development) elaborou uma classificação para a Europa, *Ranking of Excellent European Graduate Programmes*. Em 2007, apareceu o *Times Higher Education Supplement* (THES), pretendendo, como a Classificação de Shangai, abranger as universidades do mundo todo. O *Financial Times*, em 2007, classificou os 40 melhores programas europeus em administração e negócios (MBA) e, em 2008, incluiu os programas de fora da Europa. Há, também, a classificação do Centro de Estudos em Ciências e Tecnologias da Universidade de Leiden (CWTS), e um mais antigo, a do *Webometrics Rangking of World Universities*.

As disparidades classificatórias, derivadas de metodologias diferentes e resultados discrepantes, provocaram, em 2004, a criação de um grupo de peritos, sob o patrocínio do Centro Europeu da UNESCO para o ensino superior (CEPES) e o *Institute for Higher Education Policy* de Washington (IHEP). Em 2006, publicaram os critérios e princípios de classificação do *Berlin Principles on Ranking of Higher Education Institutions*, com 16 recomendações, dispostas em três categorias: os fins e objetivos das classificações, a concepção e a ponderação dos indicadores e a apresentação dos resultados (http://www.che.de/downloads/Berlin_Principles_IREG_534.pdf).

Convém reconhecer os neologismos ‘ranqueamento’ e *ranking* do ensino superior. Tornou-se um assunto recorrente da imprensa e de forte valorização pelo grande público, que extrai sua avaliação pessoal da difusão das classificações pelos meios de comunicação escrita e televisiva. É um acontecimento que veio reformular o cenário acadêmico: evidenciou a globalização do ensino superior e incitou a concorrência entre instituições e países.

A repercussão dessas classificações tem sido controversa e polêmica no meio universitário: há contestações radicais, querendo expungir da vida acadêmica essa tentação mensuradora e perniciosa ao trabalho constante e paciente do docente, tendência apresentada como nefasta pelo açulamento competitivo entre docentes. Há, também, os defensores aguerridos dessa forma de gerir o ensino superior como um *novo modelo gerencial* da universidade e dos sistemas de educação. Esse novo modelo apoia-se na *Nova Gestão Pública* (*New Public Management*), difundida nos Estados Unidos e adotada pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE), concebida para atender os desafios da globalização concorrencial dos mercados, concedendo a prestação de serviços públicos tradicionais ao mercado privado (*outsourcing*), importando do setor privado as práticas de controle financeiro e contabilístico, e adotando a responsabilização pessoal e organizacional por meio de modelos avaliativos a partir de resultados. O Chile é, constantemente, exaltado como exemplar na adoção dessa nova gestão pública, da qual deriva um conjunto de contratos público-privado típicos, cada vez mais disseminados nos Estados Unidos, como os cheques (*education vouchers*), as sociedades educacionais com fins lucrativos (*management organizations for profit, education management companies, vendors, charter schools,*), empresas que assumem parte ou o todo da escola ou somente de alguma área de gestão, mediante contratos entre o Estado e o setor privado nas áreas da educação, da saúde, do sistema prisional e tendem a se expandir para outras áreas da gestão pública. Essa teoria administrativa supõe um tipo de avaliação docente, calcada na responsabilização (*accountability*) e em indicadores de resultados. Na educação básica e nos cursos de

graduação, há um embasamento, sobretudo, nos resultados dos discentes pelo volume mensurável de produtos.

A difusão de novos processos de gerenciamento dos serviços públicos, induzidos pelo *New Public Management*, e apresentados como muito mais eficazes, pelos resultados que conseguem, tende a reconfigurar as atividades tradicionais de competência do Estado, a atenuar as fronteiras entre o público e o privado e a dar um novo desenho ao ensino superior, exigindo um determinado padrão de produtividade docente. As novas empresas educacionais tendem a dimensionar a produtividade pelo retorno do dispêndio financeiro com o docente. Isso, em muitos lugares, é feito por meio do contrato de trabalho, no qual produtividade está ancorada na hora/aula e no número de alunos atendidos pela aula. Por outro lado, esforçam-se na construção de estratégias agressivas de propaganda comercial para atrair a “clientela” estudantil, com o vitral da classificação.

A avaliação da produção docente por meio de indicadores numéricos, ainda que sob protestos veementes de viés perverso, tende a constituir um ingrediente permanente e crescente de avaliação. Para alguns, esses indicadores bibliométricos estarão nas agendas de avaliação; vieram para ficar e será difícil escapar-se deles; assim, melhor que recusá-las *in limine* é mais oportuno construir indicadores úteis, sob controle das partes e contestar indicadores mal construídos e simplistas, refutar ilações indevidas, reconhecer os limites, afastar as ciladas de um julgamento capcioso exclusivamente fundado em publicações ou resumos (PALLEZ; FIXARI; MOISDON, 2009).

Na direção de superar esses equívocos, alguns países desenvolvem o projeto europeu para avaliar os indicadores de qualidade da pesquisa, o *European Education Research Quality Indicators* (EERQI), com o objetivo de levantar novos indicadores e estabelecer metodologias de avaliação da qualidade das publicações de pesquisa no campo da educação para “melhorar a visibilidade e competitividade da pesquisa educacional européia”, que não fiquem restritos às bases de indexação e às classificações comerciais.

Outra iniciativa é o *European Reference Index for the Humanities* (ERIH), que busca estabelecer novos indicadores de avaliação da qualidade no campo das ciências humanas, e o SPRU, sediado na secretaria da Universidade de Sussex, que visa a criar uma “robusta base de dados bibliométrica para avaliar o impacto de todo tipo de resultado de pesquisa”; todos interessados em superar o monopólio crescente das bases de dados comerciais.

No Brasil, a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) desenvolveu a sua metodologia de avaliação, iniciada em 1978, para responder ao primeiro surto de expansão dos programas de pós-graduação. Em 1998, impelida pela tendência

internacional dominante e urgida pelas exigências da ampliação da pós-graduação nacional, a CAPES criou seu sistema de avaliação, adotando a avaliação trienal dos programas de pós-graduação, recorrendo, progressivamente, à classificação das fontes de informação, a fim de qualificar a produção dos programas e de seu corpo docente. A avaliação da produção científica estaria, desse modo, também apoiada na avaliação pelos pares, realizada pelos Comitês de Avaliação das respectivas áreas, que dispõem de parâmetros quantificados para suas apreciações avaliativas no *Qualis* da área de conhecimento (<http://www.capes.gov.br/avaliacao/qualis>).

O programa *Qualis* define-se como uma classificação feita pela CAPES dos veículos utilizados pelos programas de pós-graduação para a divulgação da produção intelectual de seus docentes e alunos, cujo objetivo é atender às necessidades específicas da avaliação da pós-graduação realizada por esta agência.

Nesses últimos anos, a CAPES, com o objetivo de qualificar a produção científica, aprimorou o programa *Qualis* para classificar periódicos, eventos e livros. O objetivo das classificações nas áreas é elevar o padrão de qualidade técnica e científica de produções nacionais e locais e contribuir para alcançar a excelência desejada no campo da produção qualificada dos diferentes domínios da área. Ao disponibilizar para os pesquisadores uma classificação de periódicos, eventos e livros avaliados por alguns critérios, previamente elaborados e tempestivamente divulgados, pode-se, por ilação, atribuir idêntico valor ao trabalho inserido na escala classificatória. As classificações, seus critérios e resultados (CRUZ, 2009), objeto de veementes debates, comportam outro espaço de debate.

3. ALGUMAS REFLEXÕES E SUGESTÕES

São grandes as objeções à avaliação da produtividade por indicadores numéricos ou a pertinência de se avaliar docentes por meio da quantificação de citações, principalmente, em ciências humanas e sociais.

Preliminarmente, deve-se reconhecer que a avaliação da atividade científica é um processo decorrente da vida democrática: primeiro, porque é indeclinável que todo trabalho científico auxilie no desenvolvimento econômico e social do país, pago pelo contribuinte; segundo, porque a sociedade tem que dispor de parâmetros claros dos seus dispêndios para a atividade científica e espera que sejam explícitos os critérios de prestação de contas.

A avaliação de instituições educacionais e dos docentes tem uma pretensão ambiciosa: escapar à fragmentação dispersiva, à improvisação inconseqüente, para imprimir uma racionalidade à ação e à decisão. A busca do rigor procedimental de um processo avaliativo é um suposto necessário, mas insuficiente, se não forem consideradas as implicações concretas dos pressupostos epistemológicos pertinentes que têm justificado a concepção avaliadora e os objetivos sociais e políticos relevantes que têm sido perseguidos pela atual avaliação.

A questão central é se o uso quantitativo, especialmente de citações, serve para avaliar a produção docente, a pesquisa, a docência e o pesquisador. Os indicadores numéricos, especialmente em ciências humanas e sociais, são bastante criticados pela impropriedade do parcial viés estatístico, pela ambição ingênua de exaltação de centros de excelências e pelos resultados nefastos que têm produzido.

Uma objeção ao viés estatístico é que a produção pode ser diminuta, mas muito citada; o que daria um escore alto, mas não representativo da produtividade. Para contornar esse viés, foi proposto, em 2005, o índice h (h-index), por Jorge E Hirsch, incorporado, depois, pelo ISI, Scopus e Google, a fim de compor o número de publicações com o número de citações por publicação. Assim, índice-h 5 de um autor corresponde a cinco artigos publicados que receberam ao menos cinco citações. Nas palavras do autor: o cientista tem índice h se o h de seu N_p trabalhos recebe ao menos h citações cada um, e os outros ($N_p - h$) trabalhos têm como máximo h citações cada uma. O índice contorna em parte, mas não elide o viés estatístico como têm mostrado as críticas derivadas de sua aplicação prática (COSTAS; BORDONS, 2007).

O empenho mensurador pode trazer informações elucidativas, que, no entanto, não são categóricas nem diferem, em seus resultados, das críticas que se levantam contra a avaliação pelos pares. Aprender a frequência de um evento pode ser um auxílio significativo no processo avaliativo, mas já foi sobejamente discutido, em epistemologia, a fragilidade do fundamento indutivo para generalizações, a partir de eventos singulares. Extrair probabilidades não assegura certeza, nem elimina o caráter enigmático de qualquer avaliação. Por exemplo, pode consagrar prestígio métrico e incidir em equívocos hilares: a referência pode ser repetida pelo seu caráter equivocado e ilógico, mas contribuir para elevar o nível de classificação do autor, confirmando o adágio oportunista “falem mal, mas falem de mim”. Pode, também, incitar a competição citatória para ampliar a quantidade de citações mútuas ou provocar práticas ardilosas: citações acumpliciadas, complacentes, benevolentes, obsequentes que dão um prestígio meramente métrico ao docente, ou, ao contrário, por elisão, podem

produzir deliberadamente citações omissas, arbitrárias, lacunares que dão uma posição falaz ao pesquisador. Pode, enfim, consagrar a ditadura dos indicadores, induzindo uma burocracia de controle inútil (DANON, 2010).

No que se refere ao Brasil, é inegável que o material indexado pelas grandes bases é, essencialmente, em língua inglesa (prevalentemente, norteamericana), sem esquecer que alguns periódicos publicam artigos exclusivamente sobre a realidade norteamericana. O recurso a essas bases para avaliar um docente nacional, pode mostrar-se comprometido com a hegemonia cultural dos países de língua inglesa.

A classificação, em suma, é um ingrediente na disputa pelo novo mercado docente e discente: no âmbito internacional, a competição provoca o recrutamento de docentes pesquisadores, fortemente apoiado no prestígio alcançado pelo professor, especialmente no campo da pesquisa.

O nível classificatório da universidade, alcançado pela pós-graduação, é apresentado como indicativo da qualidade da universidade. No âmbito nacional, é um expediente que tende a ser replicado com as avaliações e classificações das universidades, dos cursos de pós-graduação e de graduação. Essas classificações pretendem criar padrões de qualidade acadêmica e favorecer as escolhas aos pretendentes de ensino superior, mas são, sobretudo, usadas como fator de forte atração às demandas discentes, estimuladas por propaganda em meios de comunicação ou por folders similares a guias turísticos sobre biblioteca, laboratórios, equipamentos pedagógicos, esportivos, recreativos e condições de comodidade à vida estudantil, como restauração, alojamento, convívio, transporte, etc.

As classificações podem reforçar a dualização do ensino superior entre os centros de excelência - que figuram como vitrine do avanço científico do país e como polo atrativo de certificações acadêmicas eminentes - e as outras instituições de ensino, destinadas à provisão da mão de obra a ser formada após a educação básica.

A publicação de nível internacional pode ser enganosa: pode ter importância marginal. Uma pesquisa na área educacional, sobre o sistema de ensino brasileiro pode ter interesse limitado em um periódico internacional, mas ser absolutamente relevante para o país; por outro lado, artigos academicamente bem construídos, podem ter boa acolhida em periódico internacional, mas muito pouco interesse local.

É inegável a relevância da avaliação e a validade dos processos de aprimoramento da pós-graduação no Brasil e, conseqüentemente, a publicização dos resultados da produção acadêmica. Entretanto, não se podem ignorar, no procedimento avaliativo, as especificidades das áreas de conhecimento.

É preciso considerar epistemologicamente o processo de avaliação. As ciências da natureza e as humanas são diversas em suas práticas de pesquisa, com características metodológicas e teóricas que lhes dão a essência do que são; portanto, é preciso considerá-las também no momento de avaliar os resultados sociais, científicos, econômicos e políticos da produção do seu conhecimento. O risco da uniformização, de um padrão único, no processo de avaliação e qualificação da produção docente na pós-graduação tem provocado de discussões acaloradas quando se toma a bibliometria como indicador privilegiado da atividade e da qualidade dos programas de pós-graduação e, por derivação, dos docentes desse nível de ensino.

Os esforços feitos e os avanços conseguidos para construir um modelo avaliativo, que tenha coerência e consistência em todo o processo, constituem uma meta só alcançável com maior interlocução com os programas e pesquisadores. No entanto, apesar dos esforços da CAPES e de todos os pesquisadores, há, ainda, uma longa jornada de lutas na busca de um sistema de avaliação não só técnica e cientificamente consistente, mas também, sensível às particularidades de cada área ou cada programa

Em uma visão menos uniforme na busca de resultados de avaliação, as diferentes culturas acadêmicas poderiam ser consideradas em suas identidades e, com isso, as produções diversas de conhecimentos poderiam apresentar o seu perfil de modo mais visível e claro, social e academicamente.

É preciso, ainda, procurar caminhos que valorizem a avaliação como um processo contínuo, que gerem o bom convívio acadêmico, sem a exacerbação competitiva produtivista entre os pares, e valorizem também o ensino e a boa formação de novos docentes universitários. Uma produção científica madura requer cooperação entre pesquisadores, tempo de amadurecimento, abordagens teóricas que iluminem e sintetizem dados de pesquisa, entre outras condições fundamentais. A avaliação da produção científica sugere ter presente algumas questões recorrentes: aprofundar, de modo coletivo, técnica e cientificamente as discussões sobre avaliação da produção docente no ensino superior; continuar, buscando cada vez mais, uma aproximação da avaliação com as características da área avaliada; definir, com prazos compatíveis os critérios avaliativos; articular o processo de avaliação com processos de fomento à pesquisa a partir de necessidades sociais do país.

Essas são algumas questões que têm a intenção de fomentar as discussões sobre a avaliação da produção acadêmica docente em ciências humanas e sociais. Não se trata de um debate doméstico; está, reitera-se, no centro de um amplo debate, nacional e internacional, sobre os processos de avaliação da educação superior, sobre os critérios avaliativos e as

concepções de avaliação da produção e da qualidade das atividades científicas dos professores. Pelo fragor das provocações que levanta, retorna em todas as fases da avaliação da Capes e recrudescer nos murmúrios dos docentes da pós-graduação, no final das avaliações trienais. Os discursos contraditórios revelam os consensos e dissensos em torno do sistema atual de avaliação perseguido pela CAPES e mostram que outros caminhos são viáveis e possíveis, motivos pelos quais, apesar de crescente adesão ao processo vigente, as discussões estão longe de terminar.

REFERÊNCIAS

ACADEMIC RANKING OF WORLD UNIVERSITY – ARWU. Disponível em: www.arwu.org. Acesso em: 02 agosto de 2010.

CENTER OF HIGHER EDUCATION DEVELOPMENT. -The CHE Ranking of Excellent European Graduate Programmes. Disponível em: www.che-ranking.de/cms/?getObject=487&getLang=de – Acesso em: 28 setembro de 2009

COSTAS, Rodrigo; BORDONS, Maria. The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. **Journal of Infometrics**, v.1, nº3, p.193203.

CRUZ, Maria Lucia I, Gomes da. **Avaliação na educação superior: a contribuição da CAPES para o aprimoramento dos programas de pós-graduação stricto sensu de administração**. Tese (Doutorado). Pontifícia Universidade Católica de São Paulo: Educação: Currículo, 2009.

DALSHEIMER Nadine; DESPRÉAUX, Denis. Les classements internationaux des établissements d'enseignement supérieur. **Éducation & Formations**. 2008.

DANON, Julien. **La fièvre de l'évaluation**. Disponível: www.sciences humaines.com. Acesso em: 2 maio de 2010.

GUEDES. Vânia L. S; BORSCHIVER, Suzana. **Bibliometria: uma ferramenta estatística para a gestão da informação e do conhecimento, em sistemas de informação, de comunicação e de avaliação científica e tecnológica**. Disponível em: suzana@eq.ufjf.br. Acesso em: 24 agosto de 2009.

HARFI, Mohamed. Le classement de Shanghai et image internationale des universités : quels enjeux pour la France? La documentation française. **Horizons stratégiques**, 2006

PALLEZ, Frédérique; FIXARI, Daniel; MOISDON, Jean-Claude. **L'évaluation des chercheurs en question: 1992-2009**. Paris: Presses des Mines ParisTech, 2009.

REY, Olivier. Productivité et qualité scientifique: avec quelles publications compter? **Dossier d'actualité** n° 46 – juin-juillet 2009. Disponível em: www.inrp.fr/vst. Acesso em: 03 agosto de 2010.

SCHWARTZ, Laurent. **Pour sauver l'université**. Paris: Seuil, 1983.

Breve Currículo dos autores:

Antonio Chizzotti: Mestrado e Doutorado em Educação (PUC/SP). Pós-doutorado: Paris (INRP). Professor pesquisador Programa de Pós Graduação em Educação: Currículo da PUCSP. Linha de Pesquisa: Políticas públicas e reformas educacionais e curriculares. Pesquisa sobre os seguintes temas: epistemologia e educação; políticas públicas educacionais. Endereço eletrônico: anchizo@uol.com.br

Branca Jurema Ponce: Mestrado e Doutorado em Educação (PUC/SP). Docente do Programa de Pós Graduação em *Educação: Currículo* da PUCSP. Linha de Pesquisa: Políticas públicas e reformas educacionais e curriculares. Temas pesquisados: políticas públicas educacionais; ética no currículo. Endereço eletrônico: tresponces@uol.com.br

Aceito para publicação em 26/07/2010

