



**Avaliação:
resultados, tendências e desafios**

Luiz Nunes de Oliveira

RESUMO

Discute-se criticamente o sistema de avaliação de docentes na USP. Em função das dificuldades para avaliar o ensino e a extensão, promoções, bolsas e outras vantagens têm sido tradicionalmente distribuídas com base na análise de listas de publicações, frequentemente convertida em contabilização de artigos em revistas científicas com parâmetro de impacto acima de certos mínimos. Argumenta-se que a universidade precisa reformar seu sistema de avaliação para aprimorar a qualidade do ensino, da pesquisa e da extensão. Discutem-se princípios gerais que devem guiar os processos de avaliação e mostra-se que, a partir deles, será relativamente fácil traçar diretrizes para a reconstrução do sistema.

Palavras-chave: avaliação docente; avaliação da pesquisa; parâmetro de impacto; planejamento estratégico universitário.

ABSTRACT

A critical review of faculty evaluation at the University of São Paulo is presented. In view of the difficulties plaguing the qualitative assessment of teaching and outreach activities, promotions and distribution of fellowships and other benefits have traditionally been decided by focusing on research records, often on the number of articles published in scientific journals with impact parameters exceeding predefined minimum criteria. Here it is argued that the university needs more comprehensive, reliable evaluation methods to implement plans targeting the improvement of teaching, outreach, and research. We discuss general principles on which the evaluation processes should be based, and show that making use of them will make it considerably easy to set guidelines to rebuild the system.

Keywords: *faculty evaluation; research assessment; impact parameter; academic strategic planning.*

Não se concebe uma universidade sem avaliação. O próprio conceito de universidade deriva de um sistema primitivo de qualificação acadêmica surgido há um milênio: por volta do ano 1000, várias catedrais europeias passaram a credenciar candidatos competentes interessados em lecionar fora das instituições eclesiásticas. Com o correr do tempo, em Bolonha, Paris, Oxford e outras

idades, grupos de mestres assim licenciados se tornaram conhecidos por oferecer ensino de excepcional qualidade, e passaram eles mesmos a oferecer licenças. As universidades nasceram quando a nova modalidade de licenciamento passou a ser reconhecida. As escolas de Bolonha e Paris foram as primeiras a receber tal aval, dado pelo imperador Frederico Barbarossa. Posteriormente, várias outras universidades foram autorizadas por bulas papais ou imperiais, e logo se estabeleceu que somente o Vaticano e o Sacro Império Romano tinham autoridade para validar o licenciamento outorgado por instituições universitárias.

Tanto por isso como porque a teologia e a lei canônica tinham proeminência entre as disciplinas lecionadas, a qualidade do trabalho nas universidades medievais foi sempre controlada pela Igreja Católica. O sistema passou por uma crise quando a Reforma introduziu novos valores, em seguida modificados pela emergência da ciência moderna. Poucas universidades reagiram bem às

mudanças, e a humanidade teve de esperar muito para ver as demais se reerguerem. Por fim, no início do século XIX, as mudanças tecnológicas e sociais passaram a pedir uma nova forma de ensino superior e, guiada por forças internas e externas, a universidade respondeu. O modelo implantado por Wilhelm von Humboldt na Universidade de Berlim, apoiado em combinação de planejamento estratégico e avaliação, prosperou na Alemanha e se espalhou por boa parte do mundo. Alguns países seguiram caminhos distintos para implementar reorganizações semelhantes. Oxford e Cambridge, por exemplo, foram profundamente modificadas na segunda metade do século XIX, depois que comissões designadas pelo poder monárquico identificaram problemas estruturais.

O modelo de Humboldt chegou ao Brasil em 1934, com a criação da USP. O Estatuto de 1934, inspirado na proposta do humanista alemão, inclui dispositivos para avaliação dos docentes que, aos olhos de hoje, parecem rudimentares. A evolução da universidade, que influenciou todo o sistema acadêmico brasileiro, deu origem a um procedimento interno bem mais complexo, ao mesmo tempo em que as agências de fomento desenvolviam seus próprios métodos. Mais recentemente, diversas instituições internacionais passaram a divulgar *rankings* de universidades, que têm recebido muita atenção do grande pú-

LUIZ NUNES DE OLIVEIRA é professor do Instituto de Física de São Carlos (USP).

blico porque expressam claramente os resultados de seus levantamentos.

As universidades brasileiras têm hoje a sua disposição um elaborado sistema, com componentes internas e externas, para avaliar o próprio trabalho. Idealmente, cada uma delas exploraria seus instrumentos para medir o progresso rumo ao alvo de seu planejamento estratégico e para calibrar ações corretivas. Na prática, todavia, mesmo nas instituições acadêmicas que definem estratégias, a comunidade universitária dá muito mais atenção aos resultados das avaliações e aos ônus e bônus delas decorrentes – as bolsas de produtividade em pesquisa, por exemplo – do que às metas de longo prazo de sua instituição. Resulta que os quesitos de avaliação, e não as metas, definem o norte do trabalho acadêmico, o que nos faz perguntar se caminhamos no rumo certo.

É fácil encontrar dados que respondem ao menos parcialmente a essa questão. Já se percebeu há uma década que a produção científica brasileira avança muito mais quantitativa do que qualitativamente. Já se disse, também, que essa distorção se deve aos métodos da Capes e do CNPq, um argumento que este texto discutirá. Mais especificamente, sempre com foco na USP, veremos que se consolidou um sistema multifacetado e amplamente redundante que valoriza a taxa anual de publicações sem premiar o trabalho de qualidade superior. Na raiz desse complexo estão questões práticas que dificultam e retiram nitidez do trabalho de quem avalia a pesquisa. Os mesmos obstáculos se erguem contra a avaliação das outras atividades acadêmicas, o ensino e a extensão. Por isso, assim como não consegue aprimorar a investigação científica, a universidade não consegue melhorar o ensino ou a interação com a sociedade. Para resolver tais problemas, será necessário trabalho coordenado dos numerosos setores responsáveis pela avaliação na USP para promover uma mudança cultural que acabe afetando também o setor externo. Somente assim poderá a universidade oferecer a seus docentes condições para alcançarem a plenitude de seu potencial.

RESULTADOS

O sistema de avaliação da USP compreende vários ramos, que vão desde o estatutário até o

informal. A Tabela 1 lista os principais. As avaliações assinaladas com asteriscos dão pouca atenção às publicações em revistas científicas. Nas demais, o número de trabalhos em revistas científicas e livros tem peso importante, um critério de avaliação que é hoje empregado em todo o sistema acadêmico brasileiro. Em geral, uma comissão julgadora define um mínimo aceitável de qualidade para o trabalho – um crivo. O crivo depende da área e da subárea de concentração e da modalidade de avaliação. Em alguns casos, qualquer publicação é aceitável. Em outros, somente são considerados trabalhos publicados em revistas com parâmetro de impacto acima de determinados valores. Em algumas avaliações a contabilização é individual e pode ser feita mentalmente. Em outras, o número de trabalhos que passa pelo crivo é anotado em planilhas, acompanhado ou não por outros indicadores da atividade do mesmo candidato. Em praticamente todas, de uma forma ou de outra, contam-se os trabalhos aceitáveis.

A referência a número de trabalhos tem longa tradição. O Estatuto de 1934 exigia que cada livre-docente publicasse minimamente “um trabalho de valor” a cada cinco anos. Os números de hoje são diferentes, e o valor é especificado pelo parâmetro de impacto da revista que publica o trabalho, mas o conceito é o mesmo. Mais adiante, discutiremos a outra exigência do antigo estatuto, “realizar atividades eficientes no ensino”, mas quem conhece o cotidiano acadêmico sabe que ela caiu na sombra. Os procedimentos de avaliação instam os docentes a trabalhar com mais eficiência para publicar mais.

É por isso fácil encontrar dados para descrever os resultados desse sistema de avaliação. A Figura 1 mostra o número de artigos científicos publicados de 1995 a 2014 por três universidades de grande porte: a USP, a Universidade de Barcelona e a Rutgers State University. Barcelona fica atrás da USP na maioria dos *rankings* internacionais, enquanto Rutgers fica à frente. Os números devem ser lidos com alguma cautela, porque nem toda publicação de boa qualidade aparece em revista indexada e porque o número de revistas indexadas cresceu substancialmente no período coberto pela ilustração. Mesmo assim, o contraste entre a curva vermelha e as outras duas mostra que os docentes da USP responderam ao insistente apelo da avaliação.

TABELA 1

AVALIAÇÕES A QUE ESTÃO SUJEITOS OS DOCENTES DA USP

Processo	Órgão responsável	Natureza	Grupo avaliado
Concurso de ingresso	Unidade	li	Ingressantes na carreira
Livre-docência	Unidade	li	Doutores
Concurso para titular	Unidade	li	Associados
Promoção horizontal	CCAD e CAS	li	Doutores e associados
Relatório do período de experimentação	Departamento, unidade e Cert	li	Recém-contratados
Credenciamento para atividades simultâneas	Departamento, unidade e Cert	li	Docentes em RDIDP
Credenciamento para orientação	Comissão de pós-graduação	li	Doutores, associados e titulares
Programas das Pró-reitorias	Pró-reitorias	li ou lc	Doutores, associados e titulares*
Pedido de claro docente	Comissão de Claros	lc	Departamento*
Plano de metas	CPA	lc	Departamento
Auxílio ou bolsa	CNPq, Fapesp ou Finep	Ec	Grupo de pesquisa
Avaliação de programa	Capes	Ec	Programa de pós-graduação
Recredenciamento de curso	Conselho Estadual de Educação	Ec	Curso de graduação*
Avaliação da universidade	Conselho Estadual de Educação	Ec	Universidade*
Rankings internacionais	Times, QS, ScImago, etc.	Ec	Universidade*

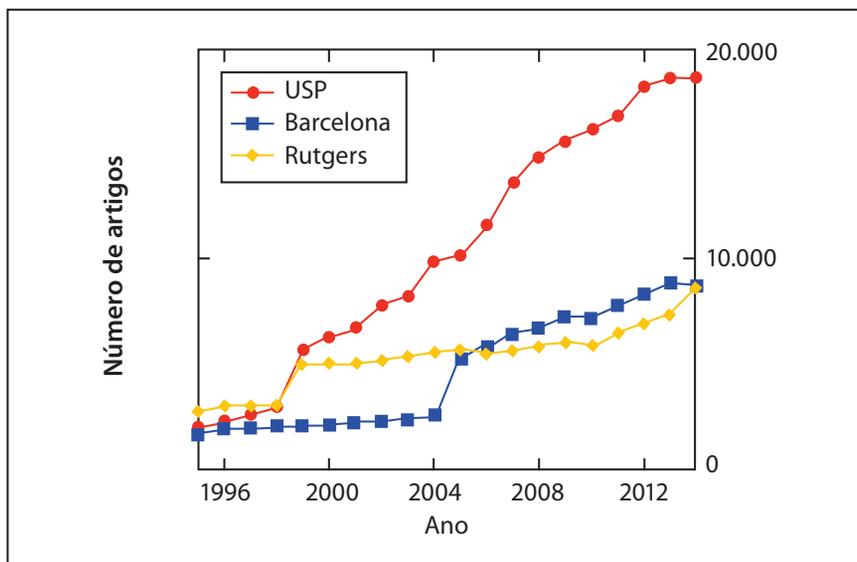
Siglas: CCAD – Comissão Central de Avaliação Docente; CAS – Comissão de Avaliação Setorial; Cert – Comissão Especial de Regime de Trabalho; CPA – Comissão Permanente de Avaliação; li – interna/individual; lc – interna/coletiva; Ec – externa coletiva; RDIDP – Regime de Dedicação Integral à Docência e à Pesquisa. Os asteriscos distinguem os processos em que as publicações em revistas científicas têm pequeno peso

Poderia argumentar-se que o número crescente de publicações espelha a evolução da ciência brasileira e que nossos pesquisadores têm publicado mais em consequência da internacionalização da universidade brasileira, mas a Figura 2 conta uma história diferente. Consideradas as publicações em revistas indexadas no período de 1992 a 2011 com pelo menos um autor da USP, Universidade de Barcelona ou Rutgers, a figura mostra o número médio de citações recebido por artigo no terceiro ano após sua publicação.

A opção pelo terceiro ano foi feita após análise das séries históricas das citações dos artigos contabilizados na Figura 1. É natural que o número anual de citações de um trabalho científico passe por crescimento inicial, à medida que ele se torna conhecido, e depois decaia, à medida que perde novidade. Nas séries históricas de citações dos artigos da USP, o máximo é alcançado três anos após a publicação; nas séries históricas da Universidade de Barcelona e Rutgers, três a cinco anos após a publicação. Para realçar

FIGURA 1

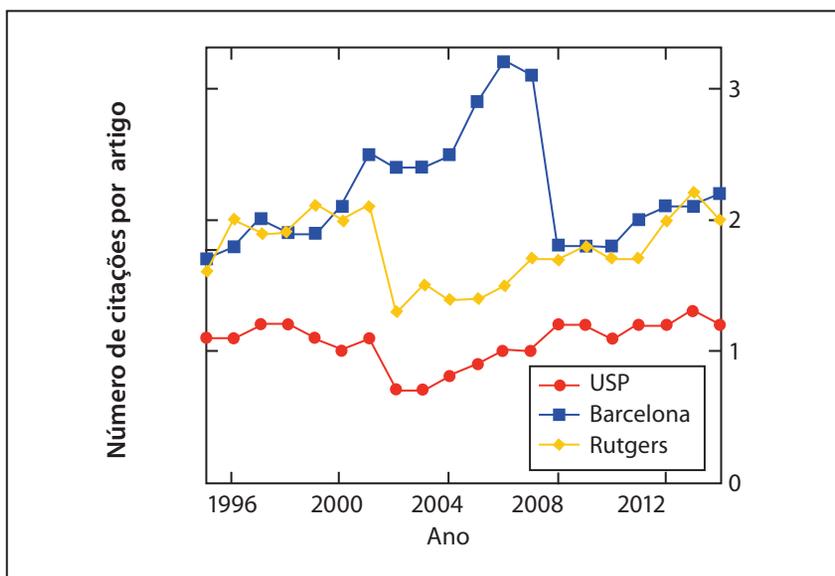
EVOLUÇÃO DA PRODUÇÃO CIENTÍFICA DE TRÊS UNIVERSIDADES NOS ÚLTIMOS 20 ANOS



Artigos científicos publicados a cada ano em revistas indexadas com ao menos um autor vinculado a USP, Universidade de Barcelona ou Rutgers
Fonte: base de dados ISI-Thomson-Reuters

FIGURA 2

EVOLUÇÃO DO NÚMERO DE CITAÇÕES RECEBIDAS POR TRABALHOS CIENTÍFICOS DAS TRÊS UNIVERSIDADES



O eixo vertical mostra o número médio de citações recebido por um artigo no terceiro ano após sua publicação
Fonte: base de dados ISI-Thomson-Reuters

os números da USP, a Figura 2 trabalha com três anos de amadurecimento.

Mesmo com esse viés favorável à universidade brasileira, o número de citações e a taxa média de crescimento do número de citações de artigos publicados pela USP são inferiores aos das universidades estrangeiras. Além disso, o gráfico associado à USP é mais estável do que os demais, uma indicação de que os autores nacionais se dedicam a áreas mais maduras, nas quais as novas descobertas atraem interesse limitado, enquanto os das outras, especialmente os de Barcelona, se arriscam mais na procura de novos veios. Esses gráficos são consistentes com a conclusão de que os pesquisadores nacionais dão menor valor à novidade e ao impacto dos resultados de pesquisa do que os pesquisadores das duas universidades do hemisfério norte. Em resumo, em contraste com estes últimos, os primeiros preferem atuar de forma a conseguir mais resultados, ainda que suas descobertas tenham menos impacto. Não se sustenta, à luz desses dados, a tese do aprimoramento da ciência nacional. Ao que tudo indica, o perfil da ciência nacional é antes definido pelo método de avaliação.

TENDÊNCIAS

A seleção por crivos traz grandes vantagens operacionais. Estabelecidos os mínimos, os currículos Lattes e as bases de dados sobre publicações científicas oferecem avaliação quase que automática de qualquer docente ou grupo de docentes. Se necessário, os mínimos podem ser calibrados para preestabelecer uma porcentagem de aprovação, e os elementos do cálculo podem ser ajustados para acomodar diferenças entre as áreas: nas tecnológicas, por exemplo, pode se convenicionar que uma patente licenciada equivale a certo número de publicações científicas. E as decisões são objetivas: conhecidas as regras, o docente objeto da avaliação nem precisa aguardar o julgamento, porque pode prever o resultado.

O avaliador que prefere uma abordagem mais estruturada se defronta com um grande desafio. Cada unidade universitária tem seu conjunto de valores, e os objetivos de curto e longo prazos variam de indivíduo para indivíduo. Comparar carreiras cultivadas em ambientes culturalmente distintos e construídas em colaboração com estudantes cujas

aspirações são igualmente heterogêneas é uma tarefa espinhosa, que requer apreciação circunspecta, raciocínio dialético, exercícios alternados de análise e síntese e conciliação de rigor com tolerância. Submeter o currículo do candidato a um crivo é uma alternativa tentadora, porque troca esse arsenal de técnicas intelectuais por uma aritmética simples, rápida e objetiva.

A simplicidade, contudo, mascara deficiências. Uma dessas se manifesta no desencontro entre as figuras 1 e 2. Métodos simples de avaliação convidam os avaliados a se concentrarem nos indicadores. Entre um problema difícil, cuja solução demandará tempo proporcional a sua importância científica, e outro mais facilmente solúvel, o pesquisador em média prefere o segundo, que, no frígir dos ovos, contribuirá mais rapidamente para incrementar o indicador relevante. A mesma lógica que conduz o avaliador ao método mais simples de avaliação faz o avaliado escolher um tema mais simples para sua pesquisa. O objeto imita o sujeito e, como os avaliadores e avaliados são membros de uma mesma comunidade, fecha-se um ciclo de realimentação que é perniciosamente eficiente, como mostra o contraste entre as posições das curvas da USP nas figuras 1 e 2.

A segunda deficiência da avaliação por crivos tem consequências mais amplas. O método não alcança as demais áreas da atividade acadêmica, porque inexistem macroindicadores para o ensino e a extensão. É verdade que alguns índices já foram propostos: as tabelas de pontuação adotadas por algumas universidades do sistema universitário federal, por exemplo, para valorizar as atividades didática e administrativa em concursos de promoção, e o peso atribuído por algumas comissões de avaliação setorial do processo de promoção horizontal da USP ao número médio de aulas semanais ministradas pelos candidatos. Assim como o volume de publicações em revistas indexadas, tais índices oferecem ao avaliador uma medida numérica da atividade do docente. Em contraste com a densidade das publicações em revistas seletivas, porém, o número de horas em sala de aula ou em um gabinete administrativo não traz nenhuma informação sobre a qualidade do trabalho do docente. Ninguém cogitaria aceitar um relatório de pesquisa que se limitasse a contabilizar o número de horas de trabalho em

laboratório, mas nos exemplos citados um procedimento análogo serviu para avaliar atividades didáticas e administrativas.

Muito embora permita valorizar a pesquisa, a avaliação por crivos faz do ensino e da extensão órfãos que dependem do entusiasmo e do arcabouço cultural de alguns docentes para progredir. Resulta progresso muito mais lento nessas duas áreas de atividade acadêmica do que na primeira e resulta também que apreciável fração dos docentes veja as salas de aula como um empecilho que a impede de viver nos laboratórios. A manter-se essa tendência, a universidade continuará a publicar mais trabalhos em revistas científicas de boa qualidade, a oferecer um ensino progressivamente menos sintonizado com as mudanças culturais e a deixar a cargo do noticiário da televisão avaliar o mérito da extensão à comunidade.

DESAFIOS

Se admitirmos como premissa que, superada uma fase de fortalecimento de seu setor de pesquisa, a universidade deve agora preparar-se para alcançar novos patamares em seus três ramos de atuação, concluiremos que ela não pode prescindir de novos métodos de avaliação. A partir do sistema de hoje, além de mudanças de caráter prático, tais como articulação entre os procedimentos listados na Tabela 1, uma reforma radical parece necessária para valorizar a pesquisa, o ensino e a extensão de qualidade e oferecer um instrumento para a universidade controlar o progresso em direção a seus objetivos estratégicos.

Não cabe aqui descrever sequer um esboço para tal reformulação. Cabe entretanto discutir os princípios que a nortearão, os quais são tão importantes quanto o esboço porque não se faz reforma sem referência a noções elementares, assim como não se ergue uma parede sem fio de prumo. Internacionalmente, reconhecem-se dois princípios que devem guiar a construção de qualquer procedimento de avaliação. O primeiro é o do *conhecimento prévio*: em qualquer momento após assinar seu contrato de trabalho, o docente deve estar informado do que se espera dele, isto é, do que será valorizado e de como serão feitas as aferições. O outro é o da *participação ativa*: o docente deve saber que é agente da avaliação,

com a responsabilidade de produzir documentos que permitam aquilatar seu trabalho.

Grosso modo, os dois princípios são respeitados pelo sistema de avaliação listado na Tabela 1, ainda que a metodologia esteja sujeita às críticas que já foram expostas. Os docentes da USP sabem que o vigor de seu trabalho será medido pelo número de publicações e portanto sabem de antemão o que se espera deles. É bem verdade que a linha divisória que define o crivo nem sempre é nítida, e que critérios inesperados são ocasionalmente introduzidos nas avaliações, mas, muito embora possam resultar em distorções graves, tais desvios são pontuais. Apenas uma pequena fração da comunidade pode queixar-se de violação do princípio do conhecimento prévio.

O outro princípio também é respeitado. Os docentes entendem que é seu dever mostrar que alcançam os requisitos mínimos exigidos pelo sistema de avaliação, isto é, produzir um número aceitável de trabalhos e depois listá-lo na seção pertinente de seu currículo Lattes. Aqui também ocorrem problemas pontuais quando os avaliadores desconsideram produções que são valorizadas por parte da comunidade. Na maioria das vezes, porém, discrepâncias dessa natureza tendem a ser rapidamente reconhecidas e logo corrigidas.

A dificuldade com a avaliação por crivos não reside na violação de princípios. O problema é ela ser incompatível com as aspirações da universidade. Precisa ser substituída por procedimentos mais eficazes, e a nova abordagem, para não constituir retrocesso, terá de honrar as duas noções basilares. Vejamos se isso é possível.

Respeitar a primeira equivale a informar o corpo docente de que a universidade está empenhada em se aprimorar, em todas as suas dimensões. O aprimoramento pede que a qualidade do trabalho de cada membro da comunidade seja incrementada. Mesmo entre aqueles que já desenvolvem um trabalho muito bom, há espaço para elevação. Os demais precisarão crescer mais. Em todos os casos, se a universidade quiser ascender a outros patamares de qualidade, a avaliação terá de focalizar o progresso em cada ramo da atividade do avaliado. No ensino, por exemplo, o trabalho de um docente será considerado bom se nesse ano tiver impacto mais positivo sobre os estudantes do que no ano passado. Quase desnecessário dizer, para

que deixe de ser um conceito meramente abstrato, o impacto precisará ser medido, o que nos conduz à segunda noção basilar.

Todo docente quer participar mais ativamente de sua própria avaliação. O interesse só não é maior porque faltam canais de comunicação. Uma abertura propiciará contribuição de dados que enriquecerá a avaliação a ponto de tornar os crivos obsoletos.

Tome-se como exemplo o ensino de graduação. Tentativas unilaterais de medir a competência didática de grupos de docentes encontram sério obstáculo na falta de informações fidedignas. Com a participação dos avaliados, diversas fontes poderiam ser exploradas. Eis alguns exemplos:

- formulários de avaliação pelos discentes, com questões formuladas pela coordenação do curso, com participação dos docentes. Tais formulários devem ser correlacionados com os índices de aprovação no Sistema Júpiter para esvaziar a queixa frequente de que as respostas dos discentes refletem a sua insatisfação com o rigor do docente;
- provas e outros instrumentos de avaliação dos estudantes, que também devem ser correlacionados com os índices de aprovação;
- pareceres de colegas, preferencialmente daqueles que ministraram disciplinas em conjunto com o avaliado;
- depoimento de (número apreciável de) ex-alunos. Isoladamente, um depoimento tem pouco valor, mas um conjunto de depoimentos coerentes constitui documento importante;
- material didático produzido pelo docente;
- documentação sobre inovações didáticas e seu impacto sobre os estudantes;

- documentação armazenada em plataforma Moodle ou equivalente;
- registro em vídeo de atividades dos estudantes.

De forma análoga, documentação oferecendo visão abrangente sobre as atividades de orientação e de cooperação com setores externos poderia ser gerada. E a participação ativa dos avaliados permitiria eliminar conhecidas limitações da avaliação da pesquisa. Por exemplo, a cópia de um artigo científico resultante de uma colaboração nem sempre permite identificar a contribuição de cada autor. Para preencher essa lacuna, bastaria que o autor da USP gerasse documentos ao longo do trabalho para descrever as responsabilidades dos membros da equipe e os discutisse nos seus relatórios.

Vemos que é relativamente fácil reformular os procedimentos de avaliação. Eleitos os objetivos da reforma, a aplicação dos dois princípios acima enunciados praticamente desenhará a planta baixa do novo sistema. Trata-se de uma tarefa relativamente simples, mas seria imprudente querer avançar mais, aqui. Desenhar a mudança não é empreitada para uma pessoa ou mesmo para um pequeno grupo, visto que sem o entusiasmo da comunidade toda proposta de alteração está destinada ao fracasso e que ninguém participa com boa vontade de um projeto que não é seu. Será necessário motivar a comunidade para a reforma. A USP precisa dela para desenvolver planos coletivos e conseguir chegar às primeiras filas dos *rankings* internacionais. Meios e oportunidade estão ao alcance de todos. Falta despertar a vontade coletiva, e aí se encontra o grande desafio, o abismo que nos separa da USP de 2034 que a sociedade deseja.

BIBLIOGRAFIA

- DE JONG, Stefan et alii. "Understanding Societal Impact Through Productive Interactions: ICT Research as a Case", in *Research Evaluation*, 23, 2014, pp. 89-102.
- ESTATUTOS da Universidade de São Paulo. São Paulo, 1934, 1969 e 1988. Disponíveis em: <http://www.usp.br/normas>.
- MAZZUCATO, Elise. "Changing Education – QA and the Shift from Teaching to Learning", in *Proceedings of the 9th European Quality Assurance Forum*. Barcelona, European, University Association, 2014.
- RASHDALL, Hastings. *Universities of Europe in the Middle Ages*. Oxford at the Clarendon Press, 1895.
- RICE UNIVERSITY SENATE. *Guidelines for Faculty Appointments, Promotion, Tenure, and Renewal of Contracts*. Houston, 2013.
- THE DANISH EVALUATION INSTITUTE. *Audit of University of Copenhagen*. Copenhagen, 2004. Disponível em: <http://www.danmark.dk/netboghandel>.
- THE PATHWAY TO THE FUTURE – *North Carolina State University Strategic Plan 2011-2020*. Disponível em: <http://info.ncsu.edu/strategic-planning/overview/pathway-to-the-future/>.
- WAY, David. *Teaching Evaluation Handbook*. Cornell University, 1997. Disponível em: <http://www.cte.cornell.edu>.