

Avaliação da produção científica: considerações sobre alguns critérios¹

Maria Helena de Almeida Freitas²

Pontifícia Universidade Católica de Campinas

Resumo

Analisa os procedimentos de avaliação da produção científica no Brasil, tanto quantitativos quanto qualitativos, principalmente no que se relaciona à avaliação por pares e às avaliações de pesquisas realizadas pelas agências governamentais de fomento. Considera que a avaliação da produção científica é realizada em todos os níveis de formação profissional dos pesquisadores, desde o ensino, mas sem haver o estabelecimento de critérios claros e gerais. Apresenta escalas de avaliação de pesquisas científicas, sendo estas consideradas como formas de estabelecer critérios racionais aos procedimentos avaliatórios, evitando a avaliação por critérios outros que não os científicos.

Palavras chaves: análise da produção científica, avaliação por pares, escalas de avaliação de pesquisas.

Some criteria for the evaluation of scientific production

Summary

The aim of this paper is to analyze both the quantitative as the qualitative evaluation procedures of the scientific production in Brazil. Mainly in what respect to the peer review and the evaluation of the researches carried out by the governmental agencies. It considers that the evaluation of the scientific production is carried out in all the professional levels of the researchers career, without the establishment of clear and general criteria is not. It presents evaluation scales of scientific researches, being these considered as forms to establish rational criteria to the procedures of analysis, preventing the evaluation for not scientific criteria.

Key words: analysis of the scientific production, peer review, scales to assessing research.

Alguns critérios para a avaliação da produção científica

O ato de avaliar, julgar e escolher como bom ou ruim, importante ou irrelevante, correto ou incorreto é um comportamento ou ação extremamente comum e habitual na vida humana, sendo ainda mais relevante na atividade científica. Le Pair (1995) afirma ser a avaliação a própria atividade científica, pois em busca de respostas aos problemas, os pesquisadores ininterruptamente avaliam conceitos, dados colhidos e teorias expostas anteriormente por seus pares.

Cientistas, assim como organizações empresariais e instituições públicas, têm buscado a avaliação como meio para assegurar o desenvolvimento e o aperfeiçoamento constantes de suas áreas, atividades e negócios.

Galembeck, em 1990, alertava sobre a impossibilidade do progresso sem a ocorrência de avaliações sistemáticas do ensino e produção científica, propondo, inclusive, a avaliação dos cursos de graduação - fato atual no Brasil com o controverso "provão".

A necessidade de avaliação da atividade científica é também motivada pelo número

¹ Parte da Dissertação de Mestrado da autora, então bolsista do CNPq, defendida na pós-Graduação de Biblioteconomia da PUC-Campinas, sob orientação de Geraldina Porto Witter.

² Endereço: Rua Monte Aprazível, 548 - Chácara da Barra. Campinas Sp, CEP 13092-64.

Telefone: (OXXI9) 236-5655 R 301 (com.). Fax: (OXXI9) 236-0470. E-mail: srbibli@prt15.mpt.gov.br.

crescente de publicações que estão sendo produzidas. Kostoff (1996) aponta a cifra de cinco mil artigos sendo produzidos por dia de trabalho no mundo todo, além de mais mil documentos de patentes, o que fundamenta, segundo o autor, uma análise computadorizada que incorpore indicadores quantitativos a fim de que se possa compreender as implicações dessa produção técnica. Tal exacerbada quantidade da produção técnico-científica fundamenta também, naturalmente, o estabelecimento de controladores de qualidade desde a criação até a publicação dessa produção. Como lembra Travassos (1998), este notável aumento de trabalhos submetidos à publicação decorre também do aumento do número de instituições mundialmente ativas atuando na área de pesquisa, o que torna difícil a filtragem.

Uma das questões aí inseridas é a antiga pressão institucional de *publicar ou perecer*, o que, segundo Spinak (1996), é fruto do sistema científico anglo-saxão, não se detectando, necessariamente, em todos os países. Mesmo assim, com a crescente e necessária ligação entre as comunidades científicas é um tanto difícil negar esse esquema, já se verificando a mesma situação no Brasil. Embora, como afirma Castro (1992: 29) as contingências de trabalho inviabilizem, muitas vezes, "...qualquer cobrança, quer por questões estruturais quer por recursos financeiros, materiais ou estando relacionado às condições de recursos humanos."

Pressão institucional, crescente socialização do ensino, uma enorme quantidade de pesquisadores vivos e produzindo, e a competitividade em relação aos financiamentos e ao poder no meio acadêmico acabam por abarrotar as bibliotecas e as bases de dados, dificultando enormemente o próprio trabalho do pesquisador, o qual vê diante de si um sem número de publicações impossíveis de serem digeridas em tempo útil, havendo hoje (mais do que nunca) o perigo sempre iminente da realização de investigações repetidas e/ou concomitantes em várias partes do mundo. Perde-se também muito tempo na escolha e leitura de material pouco válido ou de baixa qualidade - a busca da informação é uma atividade corrente em ciência, consumindo de 20 a 25% do tempo do pesquisador (Witter, 1990). O problema da quantidade de estudos e pesquisas pode ser minimizado com a publicação de bibliografias e catálogos especializados, realizada por entidades nacionais e institucionais (como CNPq, IBICT, ESAL, EMBRAP A), mas as mesmas não solucionam os problemas relativos à qualidade das pesquisas produzidas. Mesmo a atuação de pares apresenta problemas na triagem dos periódicos, o que torna a tarefa ainda mais complexa (Forum on peer reviews..., 1998).

Não que se pretenda aqui propor o controle absoluto da ciência, correndo-se o perigo de propor sistemas ou atividades autoritários de controle que fatalmente ocasionariam a perda da originalidade e da liberdade necessárias ao espírito criador. Mas também não se defende aqui qualquer tipo de *laissezfaire* científico. A avaliação da pesquisa, como expõe Bufrem (1996), é fundamental para que os órgãos de pesquisa e seus participantes possam ter uma atitude crítica e reflexiva na busca de soluções de problemas, além de "...distribuir de forma mais conseqüente os recursos para a ciência, (...) [ajudar o] norteamento do financiamento de pesquisa, (...) diagnosticar razões de crescimento desigual, diferenças de *input-output*, e até prever medidas de correção" (Meneghini, 1988: 549).

Afim de avaliar a produção e a produtividade científico-tecnológica, têm-se utilizado largamente alguns indicadores de desempenho: índice de citações, prestígio dos periódicos onde os trabalhos são *publicados*, *peer-ranking* ("onde os pares são solicitados a classificar em ordem de desempenho os grupos de área de pesquisa similar" (Meneghini & Fonseca, 1990: 631 »),

publicações em periódicos que fazem uso da avaliação por pares (o que valida o trabalho perante a comunidade científica), levantamento e índices de produtividade e levantamento quantitativo da produção por instituições, áreas do conhecimento, regiões ou países. Castro (1986) considera que o processo de avaliação de pesquisa tem duas vertentes: avaliação pelos pares ("...ancorada na reputação adquirida pelo avaliado..." (p. 28) e critérios avaliativos (a bibliometria e a ciencimétrica). Mas que "previsivelmente, as alternativas mais interessantes parecem residir em combinações dos dois métodos" (p. 28). O uso de indicadores múltiplos tem sido compreendido como o melhor modo de medir e refletir o desempenho científico também por Meneghini & Fonseca (1990) e Le Pair (1995).

As análises quantitativas bibliométricas da atividade científica partem tanto da contagem da produtividade dos pesquisadores, das áreas e sub-áreas científicas e das instituições, quanto da contagem das citações dos trabalhos dos autores ocorridas em outros textos. Essa contagem de produtividade tem a função de mapear a atividade científica e a produtividade das instituições - geralmente esses dados são utilizados por agências de fomento e órgãos voltados à política científica. São análises quantitativas que, na maioria das vezes, não enfocam questões relativas à qualidade dos trabalhos, pelo próprio aspecto escolhido para avaliar.

A análise de citações tem a pretensão de "medir" a qualidade dos trabalhos científicos via dados quantitativos, baseando-se no impacto ou importância dada pela comunidade científica a alguns autores e seus trabalhos (tomada através da contagem das citações obtidas por esses últimos).

São geralmente desconsiderados alguns fatores que podem invalidar a análise como um todo: citação por apreço aos colegas, por obediência às políticas editoriais, uso de autores consagrados para o realçamento dos textos ou simplesmente por costume de citar apenas os mais prestigiados, citação de autores da mesma instituição a fim de prestigiá-la ou a citação de autores "da moda". Em contrapartida, há a não citação de reais influências ou das fontes informais, a tradição da não-citação por parte de algumas áreas do conhecimento, a auto citação (considerada por uns e não por outros), dados tendenciosos dos índices, valorização igualitária de todas as citações (mesmo quando são citadas negativamente), autoria múltipla, homônimos, erros ortográficos, e o costume de utilizar como fontes de informações os documentos mais acessíveis (Kostoff, 1996; MacRoberts & MacRoberts, 1996; Spinak, 1996; Foresti, 1990; Motta, 1983). Em estudo de 1996, MacRoberts & MacRoberts apontaram a cifra de apenas 59% das citações a fatos científicos serem corretamente creditados aos autores de direito. Faltam estudos que, como este, dêem um contorno mais real à análise de citações.

Como todos os indicadores tomados solitariamente, as análises de citações refletem *parcialmente* o desempenho da atividade científica, não podendo *garantir* a qualidade metodológica dos trabalhos científicos, como adequadamente verificou Bruer (1985) ao comparar a frequência de citações (através do Science Citation Index) de 248 artigos de periódicos da área de educação médica continuada, com os resultados de um estudo de análise do rigor metodológico desses trabalhos (praticamente o mesmo número de citações obteve trabalhos considerados de alta qualidade como os de baixa qualidade). Concluiu que poucos artigos satisfazem todos os critérios metodológicos, havendo poucos estudos controlados, mas uma positiva e significativa correlação entre rigor metodológico e frequência de citações, *embora baixa* (p.169). Pode-se afirmar que a frequência de citações não garante a qualidade de um trabalho científico, podendo significar, como

afirmam Foresti (1990) e Spinak (1996), apenas que o trabalho em questão é mais acessível que outros.

Como procedimento de indicação da qualidade de trabalhos científicos, tem-se também a avaliação por pares (*peer-review*), que é a avaliação de trabalhos científicos realizada por consultores, os quais são pesquisadores mais experientes ou especialistas de áreas específicas, sendo o tipo de avaliação mais utilizado pela comunidade científica nacional e internacional. A avaliação por pares é utilizada na análise de artigos enviados para publicação nos periódicos científicos, na avaliação do impacto da pesquisa (Kostoff, 1997), por agências de financiamento de projetos (FINEP, CAPES, FAPESP) na escolha de pesquisas que receberão verbas, na delimitação da quantidade de bolsas de estudos enviadas às instituições de ensino com diferentes graus de poder de decisão, e até em decisões quanto a promoções nas carreiras acadêmicas. As avaliações das pesquisas ocorrem, geralmente, de duas maneiras: "... revisão via correio por revisores *ad hoc* e revisão via painel - *ad hoc* ou não -, normalmente subsidiada por revisores *ad hoc* via correio (...) [as quais] têm os caráter de recomendação. Quanto à decisão final, a aprovação da proposta cabe a outra instância da agência, ao nível departamental e/ou de direção superior" (Mello, 1997: 3-4). Some-se a esses procedimentos o processo de visitas aos solicitantes de financiamento, realizadas por pares da comunidade e técnicos das agências de fomento, processo avaliativo realizado, no Brasil, pelas instituições CAPES e FINEP.

A avaliação por pares baseia-se no pressuposto de que dado o treinamento, o conhecimento especializado e a vasta experiência com a atividade de pesquisa, os pesquisadores chamados a avaliar "...são os mais bem aparelhados, capacitados para julgar as intenções manifestadas nas propostas, sendo a revisão capaz de distinguir propostas com idéias bem formuladas, inovadoras, com mérito, ao invés daquelas conservadoras, imperfeitas, espúrias..."(Mello, 1997: 5).

Meneghini & Fonseca (1990) consideram a utilização da avaliação por pares uma questão problemática quando se trata de áreas relativamente pequenas e com poucos recursos, onde avaliadores e avaliados competem pela mesma verba - pois a avaliação por pares deve ser uma avaliação desinteressada, competente e independente. Kostoff (1997) também aponta problemas na utilização da avaliação por pares: parcialidade dos avaliadores causada por razões não técnicas; o "efeito halo", o que favorece sempre os mesmos pesquisadores, temas e instituições, e os diferentes critérios na avaliação e interpretação. Além do mais, há que se considerar o considerável custo necessário para a manutenção de um sistema organizado e sério de avaliação por pares (Kostoff, 1997: n.p.).

Entretanto, a despeito de suas limitações (Forum on peerreviews..., 1998), esta avaliação é a mais recomendável para a seleção de textos para periódicos, esperando-se que os pares aperfeiçoem-se, superem-se e tenham sempre presente que a avaliação deve ter por "alvo a justiça, a excelência e a originalidade". Qualquer outra consideração discriminatória é "antiética e inaceitável" (Travassos, 1998: 8).

Assim como a análise de citações, comentada antes, a avaliação por pares requer a utilização de outros indicadores de desempenho na maioria das situações,~ a fim de complementar os mecanismos de avaliação existentes. A utilização da avaliação por pares, de forma solitária, é própria da análise dos trabalhos para publicação. De qualquer maneira que seja utilizada, a avaliação por pares deveria seguir algumas normas básicas como a ética, o apego à cientificidade e à criticidade acima de tudo, e o fato do avaliador não competir com o avaliado por benefícios

financeiros ou sociais.

A falha no sistema de avaliação pode ser retratada pela situação protagonizada pelo físico norte-americano Alan Sokal que, informado com o fato do periódico da área de ciências sociais "Social Text" ter-se negado a publicar um artigo seu, escreveu outro, no estilo da própria revista. O segundo artigo no entanto, era "um enroladíssimo ensaio (u.), não mais que uma piada repleta de frases sem sentido, para dar a impressão de que estava questionando a validade da mensuração da realidade física" (Campos, 1996: np). Logo depois, Sokal publica em outro periódico, o artigo original, ridicularizando os editores e, conseqüentemente, o sistema de avaliação do primeiro periódico. Por que o segundo artigo de Sokal foi publicado? Talvez porque Sokal seja um físico respeitado na comunidade científica americana e que no texto defendia, embora de uma maneira confusa, as crenças dos avaliadores. Considerando-se que a avaliação sistemática de pesquisas é fato extremamente recente na comunidade científica, esse tipo de acontecimento pode ser utilizado como exemplo a ser evitado, tanto para a otimização quanto para a própria validação do sistema de avaliação como um todo.

Avaliação da produção científica no Brasil

A produção científica de um país pode ser avaliada em várias instâncias e de variados modos: através do cotidiano das universidades ou empresas, análise dos produtos da atividade científica (patentes, inventos, relatórios, artigos, dissertações e teses etc.), aceitação de trabalhos para encontros científicos ou para publicação (o que pressupõe análise de editores ou pares), análises quantitativas de produtividade e de impacto (via citações), relação de patentes solicitadas, verificação de novos produtos tecnológicos enviados ao mercado, análise das citações de patentes, até avaliações formais e institucionalizadas realizadas por entidades científicas e/ou governamentais.

A região Sudeste concentra "54% de todo o pessoal envolvido com C&T, 73% de todos os doutores, cerca de 70% de todos os grupos ativos de pesquisa, 73% de todos os alunos de mestrado e 92% de todos os doutorandos" (Guimarães, 1992) - de 60 a 75% dos financiamentos são concentrados na região.

O país oferece 1.159 cursos de mestrado e 616 de doutorado (Perfil da Pós-Graduação, 1998: n.p.), que são mantidos, quase em sua totalidade, pelo dinheiro público através das universidades públicas, do financiamento de projetos e da concessão de bolsas de estudo, que são repassados dos governos estaduais e federal por algumas agências (FINEP, CAPES, CNPq), o Programa P ADCT e as agências estaduais, dentre as quais é de se destacar a F APESP, a mais antiga dentre as atuantes.

As quatro agências e o programa citado, além de serem os financiadores de quase a totalidade da atividade científica no Brasil, são os realizadores dos principais sistemas de avaliação da pesquisa em C&T, um sistema plural e nacional de financiamento da atividade científico-tecnológica não poderia prescindir de um sistema de avaliação dos projetos que financia, sob pena de desperdiçar recursos, normalmente escassos e irregulares. Todas essas avaliações são realizadas com a presença de pareceristas, dividindo-se esses em pesquisadores (membros convidados ou eleitos pela comunidade científica) e técnicos das agências.

O Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) foram criados em 1951, a Fundação de Amparo à Pesquisa no Estado de São Paulo (F APESP), em 1960 (inicialmente com a

finalidade de apoiar a pesquisa da USP) e a Financiadora de Estudos e Projetos (FINEP), em 1967. O PADCT, programa inter-agências financiado pelo governo brasileiro e pelo Banco Mundial, iniciou suas atividades em 1984.

O Programa de Apoio ao Desenvolvimento Científico e Tecnológico (PADCT) é fruto da parceria entre o governo brasileiro e o Banco Mundial (BIRD), tendo sido criado para atuar como um instrumento complementar à política nacional de fomento à C&T, com vistas ao aumento quantitativo do apoio financeiro à pesquisa, e a introdução de novos critérios, mecanismos e procedimentos, além da definição de áreas prioritárias.

A secretaria-executiva do PADCT é formada por membros do Ministério de Ciência e Tecnologia e das três principais agências federais (CNPq, FINEP e CAPES - encarregadas dos atos administrativos, trâmites das propostas, liberação dos recursos e contratação e acompanhamento dos projetos). Membros da comunidade científica, tecnológica e empresarial e representantes da Secretaria Executiva e das agências formam os grupos e comissões que definirão políticas, formularão o planejamento e analisarão as propostas encaminhadas ao órgão.

O processo de financiamento dos projetos no PADCT inicia-se com a preparação de planos anuais realizados pela Secretaria Executiva junto com os coordenadores dos grupos de trabalho. Esses planos apresentam a disponibilidade de recursos a cada subprograma. A partir deles os Grupos Técnicos elaboram editais (destinados ao público alvo) contendo informações básicas e os tipos de projetos e temas que serão aceitos. Cada agência é responsável por determinados subprogramas - recebe as propostas a elas relacionadas analisando sua adequação com os requisitos publicados nos editais, configurando-se como o primeiro filtro do processo de avaliação. A partir daí são escolhidos por representantes das agências e coordenadores, os nomes que comporão os comitês assessores e os consultores *ad hoc* (dois para cada proposta).

Via correio são enviadas as propostas aos consultores - estima-se o retorno de 85% dos pareceres (30% das propostas seguem os trâmites com apenas um parecer, havendo propostas - uma pequena parcela - que não recebem nenhum parecer, o que dificulta os encaminhamentos seguintes).

Mello (1997) afirma que os pareceres *ad hoc* tem sido considerados como de baixa qualidade pelos técnicos do Programa. Mas, mesmo assim, não se elabora um banco de nomes organizado por subprogramas, com endereços atualizados etc.; além disso, "não há um controle de qualidade sobre os pareceres dos *ad hoc*, [e] nada se registra quanto ao profissionalismo dos consultores. Deveria se evitar a indicação daqueles pareceres baixa qualidade" (Mello, 1997: 11). Entretanto, para que isto fosse efetivado seria necessária uma pesquisa de avaliação dos pareceres.

Mesmo com esses problemas o sistema de avaliação do PADCT parece ser mais criterioso que os das agências de fomento, pois logo com o retorno dos pareceres, os grupos são reunidos para uma discussão em torno das exigências e critérios expostos nos editais, os quais serão os parâmetros indicadores dos julgamentos das propostas, além das orientações da Secretaria Executiva. Os projetos são apreciados item a item, debatidos e comparados com o especificado nos editais. Após as discussões, cada projeto recebe um parecer preliminar do relator principal e notas (por item) de cada membro do comitê. Após a análise de todos os projetos, esses são classificados de acordo com as notas que receberam dos membros, havendo a escolha daqueles que obtiveram uma pontuação superior a um mínimo estabelecido.

O Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, vinculado ao

Ministério da Ciência e Tecnologia, é uma das principais agências governamentais de fomento. Realiza três atividades básicas: fomento, execução de pesquisa e informação e difusão de ciência e tecnologia. A principal ação desenvolvida pelo CNPq é o fomento, sendo dirigido essencialmente para a formação de recursos humanos e apoio às pesquisas, através de programas de bolsas (no país e no exterior), bem como apoio financeiro a cursos e projetos tecnológicos.

Até por volta de 1970, as avaliações das propostas enviadas ao CNPq eram realizadas informalmente por "um grupo de cientistas transformados em administradores" (Castro: 1985: 32). A partir desta época, tendo havido uma reformulação interna, foram criados comitês assessores por áreas do conhecimento para realizarem a atividade de avaliação das propostas.

O CNPq conta hoje, com a participação de pesquisadores, tecnólogos, empresários e políticos em seus diversos órgãos colegiados, sendo os mais importantes o Conselho Deliberativo e os Comitês Assessores. O Conselho Deliberativo é o órgão que elabora as diretrizes de atuação do CNPq. Já os Comitês Assessores, estrutura do sistema de avaliação da agência, atuam na concessão de bolsas, auxílios e julgamento das solicitações. Os consultores têm mandato bienal, reunindo-se duas vezes ao ano para analisar propostas enviadas de todo o Brasil. A rotatividade dos consultores pode evitar, de certa forma, a formação de grupos de interesse. Os trâmites das propostas, verificação de documentos e questões de prazo são administrados pelo corpo técnico do CNPq.

Críticas têm sido feitas à utilização dos recursos da agência, principalmente em relação à utilização de representativa fatia do orçamento na manutenção do próprio órgão, e quanto ao fato dos recursos serem divididos entre os comitês assessores e os programas internos. Como Castro (1985: 33) afirma, os "funcionários da instituição disputam com os assessores da comunidade a administração dos recursos do órgão."

A região geográfica de origem desses assessores é, majoritariamente, a Sudeste (60%). Já institucionalmente, 86% deles originam-se de universidades (e não de institutos de pesquisa ou da iniciativa privada), embora na escolha dos pareceristas procure-se fazer ajustes que equilibrem as diferenças regionais e institucionais tão comuns no País.

Em relação ao processo de avaliação, Guimarães (1992) considera o CNPq como a agência que comporta a maior penetração da comunidade científica entre as agências consideradas.

A FINEP é uma empresa pública vinculada ao Ministério da Ciência e Tecnologia, tendo como objetivo o financiamento de estudos, pesquisas, projetos e programas que visem o desenvolvimento sócio-econômico, científico e tecnológico do país.

Atualmente, a FINEP tem aceito pedidos de financiamento nas seguintes áreas: apoio ao desenvolvimento científico e tecnológico, gestão da qualidade, apoio a seminários e eventos, pré-investimento, apoio às empresas de setores estratégicos e de base tecnológica, parceria universidade-empresa com incentivo, capacitação de empresas em gestão ambiental, apoio à participação de empresas em feiras e eventos tecnológicos, apoio tecnológico às micros e pequenas empresas e educação para competitividade.

Seu sistema de avaliação utiliza predominantemente pessoal técnico próprio, dispondo de amplo quadro de pessoal recrutado na comunidade acadêmica. A comunidade científica participa *ad-hoc*, e com intensidade bem menor que nas outras agências. A penetração da comunidade científica no sistema de avaliação da FINEP tem oscilado na mesma proporção que a situação orçamentária da agência, e na proporção inversa ao aumento da ênfase política de fomento à pesquisa tecnológica.

O sistema de avaliação da FINEP mostra-se, pois, muito mais centralizador que os das outras agências, o que pode facilitar escolhas e avaliações de cunho político (a participação da comunidade científica também tem a função de controle da atividade das agências). Ao mesmo tempo, esse tipo de sistema minimiza um problema comum em relação à avaliação por pares, que é a concorrência pelo mesmo recurso, a qual pode ocasionar algum tipo de incorreção na análise dos projetos.

Na análise das propostas são utilizados dois ou três consultores, podendo ocorrer também, conforme o departamento, visitas à instituição, havendo o acompanhamento de membros da comunidade científica para uma avaliação do mérito e da factibilidade do projeto. Pode ainda haver visitas de acompanhamento, mas não se mostra uma prática muito disseminada no sistema como um todo.

Não são apresentados os critérios que regulam o mérito dos projetos - são comuns as mudanças de normas e programas a cada administração. Dessa maneira, aliando-se a ausência de uma norma institucional formalizada à inexistência de um diretório atualizado das atividades científico-tecnológicas, torna-se muito mais fácil resultar avaliações de cunho subjetivo. É interessante observar que na década de 70 a FINEP utilizava estudos mapeadores como parâmetro para as avaliações. Mas, nos dias de hoje já não é prática comum, somente havendo a utilização de um diretório (da CAPES) em propostas que envolvam cursos de pós-graduação senso estrito.

A F APESP é considerada a mais importante e mais atuante agência estadual do país. Atua, como todas as outras, em todas as áreas da Ciência e da Tecnologia, financiando bolsas, auxílios para projetos de pesquisa e atividades variadas de apoio à investigação científica. Foi criada em 1960 para atender as necessidades da USP, então a única universidade estadual do Estado de São Paulo, no sentido de ser um organismo autônomo de apoio à pesquisa, mais eficiente e ágil nas decisões que, costumeiramente, a administração pública, sendo gerido por especialistas qualificados. Seu orçamento é baseado na transferência de 1 % do total da receita tributária do Estado de São Paulo.

A Fundação apóia propostas apresentadas pelos pesquisadores do Estado de São Paulo, baseando as decisões em função do mérito das mesmas, tendo sido avaliadas por uma assessoria científico-tecnológica. Somente no ano de 1996 foram analisados 13 mil pedidos de apoio financeiro.

A FAPESP é administrada por um Conselho Superior que elabora suas orientações gerais e decisões quanto à política científica, administrativa e patrimonial. A metade dos seus membros é escolhida pelo governador do Estado, três são escolhidos pela USP e os outros três pelas instituições de ensino e pesquisa do estado. Há também um Conselho Técnico-administrativo que escolhe os membros, dirige e orienta as 12 comissões (divididas por área de conhecimento), as quais têm a função de avaliar as propostas enviadas.

A avaliação do mérito científico ou tecnológico das propostas (pedidos de bolsas de formação, auxílio a pesquisas, realização de congressos, participação em congressos etc.) é realizada por assessores escolhidos na comunidade científica, de acordo com a área do projeto a ser avaliado. Segundo Guimarães (1992), o sistema de avaliação da F APESP é baseado na avaliação por pares, onde o predomínio de razões políticas sobre as de mérito é praticamente inexistente, o que nem sempre ocorre nas agências federais. As propostas enviadas ao órgão são categorizadas (pelo próprio solicitante) em uma das doze áreas de conhecimento estabelecidas pela F APESP,

sendo distribuídas pelo Conselho Técnico-administrativo às doze comissões correspondentes, para análise. Cada comissão escolhe os pesquisadores que avaliarão as propostas, cabendo-lhe apenas encaminhá-las, sem emitir julgamentos de mérito. Com o retomo dos pareceres, a comissão pode acatá-los ou solicitar outros. As propostas aceitas (parecer positivo) são encaminhadas pela diretoria científica a outro grupo de pares para uma segunda análise, tendo esse procedimento uma importante função: "Essa instância tem a tarefa de compatibilizar critérios eventualmente diferentes utilizados pelas várias coordenações englobadas numa mesma grande área. Com isto, procura-se minimizar eventuais incoerências de julgamento decorrentes dos perfis de demanda em cada coordenação"(Guimarães, 1992, p. 24-25).

A partir daí, a avaliação da proposta está concluída. Já no caso de auxílios temáticos de equipe, a avaliação passa por três pareceristas (um pesquisador do Estado de São Paulo, um de outro estado e um terceiro pesquisador que esteja trabalhando no exterior).

A CAPES foi formada com o objetivo de realizar todas as ações necessárias ao aperfeiçoamento do pessoal do ensino superior, acabou por ter um papel decisivo na formação e estruturação de toda a pós-graduação brasileira, no que se refere aos cursos senso estrito.

Como todas as agências, o sistema de avaliação da CAPES é bastante recente, tendo se estruturado nos anos de 1976-77. Os cursos de mestrado e doutorado brasileiros são o público-alvo dos programas de financiamento da CAPES e o objeto de análise das avaliações. A agência avalia com regularidade, praticamente todos os programas de pós-graduação do país, a partir dos seguintes procedimentos: envio de relatórios dos cursos através das próreitorias (ou similares em universidades e institutos de pesquisa), discussão e debates acerca dos relatórios enviados por comissões de pares, as quais conferirão um grau de desempenho aos programas. Visitas da CAPES aos cursos é um procedimento comum e têm sido realizadas periodicamente.

As avaliações da CAPES, antes sigilosas, tomaram-se transparentes - os graus conferidos aos programas de pós-graduação brasileiros vêm sendo publicados nos maiores jornais do país, causando impacto social, o que acabou por valorizar o trabalho da agência. Outra característica desse sistema é a utilização crescente de indicadores qualitativos, a uniformidade de critérios baseados em comparações intra-áreas.

É interessante notar que, nos últimos 17 anos, houve um significativo aumento dos graus A e B (o que, hoje, seriam 6 e 7 para programas com Doutorado e 4 e 5 para programas com Mestrado apenas). Spagnolo (1994) comenta que se fossem considerados somente os graus A e B conferidos, o Brasil possuiria um dos melhores sistemas de pós-graduação do mundo. A questão é que a CAPES, em suas avaliações ao longo dessas duas décadas, havia assumido um papel muito mais pedagógico e de formação do que pura e simplesmente de avaliador, "mostrando quais as características desejáveis de um curso em termos de corpo docente, atividades de ensino, pesquisa, orientação, infra-estrutura e os demais aspectos relacionados à atividade de pós-graduação" (Spagnolo, 1994: 12).

Havendo uma pós-graduação mais estruturada no país, e já tendo se configurado o momento da CAPES rever seu sistema de avaliação, a agência mudou, em 1998, os seus critérios de avaliação, modificando o seu papel de pedagógico para o de avaliador, de forma a se chegar a um padrão de excelência internacional. O sistema de avaliação passou a ter como foco e objetivo a análise da capacidade de formação dos recursos humanos dos programas de pós-graduação e não mais o desempenho individual dos cursos. A escala de avaliação passou a ser de 1 a 7 (até 5 para

programas com Mestrado e até 7 para programas com Doutorado), sendo baseada nos seguintes critérios: modelo flexível de pós-graduação, relação entre o mestrado e o doutorado, integração do programa com a graduação, e a relação entre o projeto do programa e o impacto de sua atuação para a instituição e para a sociedade, entre outros. (Coordenação de Aperfeiçoamento de..., 1998)

Os sistemas de avaliação descritos têm a finalidade de orientar o processo de tomada de decisão no financiamento de projetos, sendo baseados na avaliação por pares. Excetuando o sistema de avaliação do PADCT, somente os critérios mais amplos são definidos, como temas relevantes, clareza da proposta e do método, curriculum dos orientadores (no caso de atribuição de bolsas de estudo) ou dos responsáveis pelo projetos ou equipes de trabalho (no caso das avaliações dos *cursos* pela CAPES, os critérios são bem rigorosos e claros - mas não é estabelecido nenhum controle de qualidade específico dos produtos de seus bolsistas). Os critérios específicos de mérito e qualidade utilizados no momento de avaliação pelos pareceristas são, conservadas as devidas proporções, praticamente pessoais, ou seja, não são claramente estabelecidos pelas agências, podendo variar de pesquisador para pesquisador ou de técnico para técnico, o que pode invalidar o processo como um todo.

A escolha dos avaliadores e consultores varia um pouco de agência para agência, mas, de uma forma geral, os nomes são retirados de indicações de universidades de renome, de outras agências de fomento e de institutos de pesquisa. No PADCT o processo de escolha dos representantes da comunidade científica é mais centralizado que nas agências.

A pós-graduação brasileira é um fenômeno extremamente recente, tendo começado, na maioria das universidades, na década de 60 e se institucionalizado na década de 70. Em pouco mais de três décadas, passou de 100 cursos para quase 1.800 cursos estrito sensu, sendo, conforme já referido, 1.159 cursos de mestrado e 616 de doutorado. Dentro do sistema educacional brasileiro, a pós-graduação tem sido considerada o nível de melhor performance e de maior êxito.

Também recente é a avaliação da produção científica (e a relação avaliação-financiamento), tanto em nível nacional quanto internacional, tendo se iniciado na década de 60 quando os recursos para a C&T começaram a diminuir.

Portanto, pode-se considerar que o sistema de avaliação da pesquisa científica no Brasil ainda situa-se, historicamente, em seu início. Necessário é que seja mantido e estruturado sobre procedimentos os mais democráticos e abertos possíveis, além de técnico, a fim de que o mérito científico seja premiado, a comunidade científica possa participar e opinar, mantendo-se segura quanto aos critérios avaliativos, os recursos sejam distribuídos e não concentrados e a sociedade tenha dispositivos sérios de controle das atividades que, afinal, financia.

Avaliação e qualidade de dissertações e teses

Ao se deparar com um estudo, um artigo, uma obra de cunho científico, o pesquisador naturalmente analisa alguns pontos ou partes do trabalho. A primeira questão a ser considerada é o fato do pesquisador ter escolhido aquele determinado trabalho para ler ou estudar: o título chamou sua atenção, foi produto de uma busca, versa sobre algo que o interessa, algo relevante no momento ou algo que tem sido alvo de discussões em seu meio (Witter, 1997). Então, o primeiro ponto a ser considerado é o assunto de que trata o trabalho. O assunto deve ser relevante quer para o pesquisador, quer para o meio científico, quer ainda para a sociedade.

Ao prosseguir em sua análise do trabalho (que já passou pelo 1.0 crivo avaliativo - assunto relevante) o pesquisador verifica em suas primeiras páginas se o trabalho trata, realmente, do que está intitulado. Após isso, enfoca as citações e referências, procurando verificar se os trabalhos mais conhecidos, relevantes e básicos daquele assunto foram usados como fonte de informação. Quanto mais o pesquisador conhece do assunto mais terá este comportamento, pois um trabalho que não utilize fontes fidedignas básicas (consideradas como tais pelo meio científico específico) tenderá a repetir dados, informações e resultados, ou basear-se-á em fórmulas antigas ou ultrapassadas.

Ao examinar os dados e as conclusões, o pesquisador pergunta-se como o autor do trabalho chegou àqueles resultados. Assim, passa a examinar o método e as técnicas utilizadas.

Dependendo da avaliação que o pesquisador der ao trabalho, passará a lê-lo atentamente, página por página e ordenadamente, ou o deixará na estante ou mesa, indo percorrer o acervo em busca de algo mais interessante, relevante, inovador e criterioso. E tudo isso é feito em alguns poucos minutos.

A partir deste exemplo pode-se esquematizar uma avaliação de pesquisa: (1) assunto relevante; (2) objetividade e condução lógica; (3) fontes utilizadas (bibliografia) e (4) métodos e técnicas desenvolvidos e (4) discussão e conclusões.

No caso de dissertações de mestrado e teses de doutorado há vários avaliadores: o primeiro a avaliar a dissertação é o próprio produtor e, em seguida, o seu orientador. O crivo de sua avaliação deste último reflete o adotado pela sociedade e instituições que o formaram e que o contrataram para realizar esta função, entre outras. Ou seja, diante daquela instituição o orientador é absolutamente capaz de avaliar o trabalho do aluno.

No Brasil, a segunda etapa de avaliação do trabalho do aluno de pós-graduação é o chamado exame de qualificação, onde o aluno apresenta seu projeto de dissertação a outros professores doutores, aí obtendo uma análise do trabalho proposto e/ou em realização. A partir daí, ele pode dar continuidade à pesquisa projetada com as propostas e mudanças sugeridas pelos professores no exame.

Sendo avaliado constantemente por seu orientador, o aluno segue até a terceira etapa avaliativa: a defesa de sua dissertação/tese, onde oralmente defenderá o trabalho que realizou.

Permeando todas essas etapas estão os órgãos controladores e de fomento dos cursos e das universidades. O aluno para participar da pós-graduação deve preencher requisitos previamente estabelecidos - requisitos legais, educacionais e normas da instituição. Da mesma forma os cursos, para funcionarem, devem preencher outros requisitos e normas legais e oriundos das políticas educacionais do momento. Além do que, para que os cursos obtenham recursos governamentais - fundamentais no Brasil para a existência dos cursos de pós-graduação - devem, além de terem preenchido todas as exigências, apresentar um mínimo dos resultados esperados.

Vê-se que todos os momentos da pesquisa científica, desde a primeira fase, da formação do pesquisador até a sua prática, estão sujeitos a avaliações periódicas e sistemáticas, subliminares ou não, corretas ou não.

A qualidade de uma pesquisa, estudo ou trabalho científico não é algo dado ou inerente ao que se chama de "Ciência". É regida, basicamente, pelo meio científico, ou seja, pelo conhecimento acumulado e organizado [historicamente] pelas instituições científicas e educacionais, as políticas nacionais e as atividades e cultura dos próprios cientistas. Um dos fatores essenciais do que se

chama "qualidade" é o desenvolvimento histórico da área científica determinada. O que há vinte anos poderia ser considerado um trabalho fundamental, importante e de qualidade, hoje pode ser considerado plágio, por melhor que seja a estrutura metodológica em que esteja apoiado.

Eco (1989: 21-23), ao discutir cientificidade, acaba definindo os requisitos básicos a um trabalho: "1) O estudo debruça-se sobre um *objeto reconhecível e definido de tal maneira que seja reconhecível igualmente pelos outros*. (...); 2) O estudo deve dizer do objeto *algo que ainda não foi dito* ou rever sob uma óptica diferente o que já se disse. (...); 3) O estudo *deve ser útil aos demais*. (...); 4) O estudo *deve fornecer elementos para a verificação e a contestação das hipóteses apresentadas* e, portanto, para uma continuidade pública. Esse é um requisito fundamental." (Grifos do autor)

A importância do tema é entendida como requisito a um bom trabalho por praticamente todos os autores consultados. Como Eco, Luma (1991 :26) define pesquisa científica como "uma atividade de investigação capaz de oferecer um conhecimento novo a respeito de uma área ou de um fenômeno, sistematizando-o em relação ao que já se sabe a respeito dela", sendo seus requisitos "a existência de uma pergunta que se deseja responder; a elaboração (e sua descrição) de um conjunto de passos que permita obter a informação necessária para respondê-la; [e] a indicação do grau de confiabilidade na resposta obtida" (Luma, 1991: 27).

Bruer (1985), em artigo em que analisa a relação entre rigor metodológico e frequência de citações em literatura médica, utiliza-se dos seguintes padrões de análise para determinar o grau de rigor metodológico e, por conseguinte, da qualidade da pesquisa científica: descrições suficientes do tema e tópicos da pesquisa e da intervenção do pesquisador, a fim de permitir relações e generalizações; maior pontuação para experimentação controlada com amostra aleatória e menor para estudos do tipo antes-e-depois; o grau de modificação do ambiente ou sujeito ocasionado pela intervenção do pesquisador; e a significância do trabalho que, no relato, relaciona a intervenção e os resultados da avaliação (resultados estatísticos).

Dissertações e teses não fogem às regras dos trabalhos científicos, pois, como exposto por Lakatos & Marconi (1995), a dissertação é um tipo de trabalho científico realizado para obtenção do título de mestre, ao que acrescenta que a tese visa a obtenção do doutorado. Mas representando mais do que isso, dissertações e teses fazem parte da literatura cinzenta, isto é, literatura científica e/ou tecnológica oriunda diretamente de pesquisas e atividades recentes finalizadas ou em execução, metodologias mais controladas, reduzida tiragem, divulgação restrita e difícil acesso, a qual, por algumas dessas características, provê os pesquisadores senão na maior parte de suas necessidades, como afirmam Población, Noronha & Currás (1996: 228), então em boa parte das necessidades, pois tal produção tem caráter didático (treinamento), muitas vezes natureza reflexiva, sistematização, ordenação e interpretação dos dados. A literatura cinzenta, dela fazendo parte as teses e as dissertações, relata a pesquisa inédita e a atual, não podendo ser desprezada pelos pesquisadores.

Como um tipo de trabalho científico, a dissertação não tem obrigatoriamente que discorrer sobre um tema original, desde que proponha uma visão original do mesmo objeto. Como expõe Fazenda (1991), um tema muito pesquisado, embora possa ser considerado desalentador, não deixa de ser relevante à ciência, pois novas formas de investigação podem revelar aspectos desconhecidos do problema. Ao que se pode acrescentar a relevância da réplica em ciências.

O trabalho científico deve, acima de tudo, acrescentar. Dizer o não dito, ser útil científica ou

socialmente, mas de forma que assim seja reconhecido pela comunidade científica. No entanto, nem sempre a comunidade científica pode ser considerada um bom termômetro, pois nenhum grupo pode dar a última palavra em relação às certezas científicas. De qualquer maneira, dentro das devidas proporções (políticas e relações de poder à parte), os pares podem ser considerados avaliadores competentes dos trabalhos científicos.

Assim como a significância e relativa originalidade do tema, o bom trabalho científico deve relatar farta e claramente os procedimentos e passos utilizados para se chegar às conclusões.

A análise da qualidade de dissertações e teses não é realizada por nenhuma instituição de fomento no Brasil, ficando essa função para as instituições de ensino superior no decorrer dos cursos de pós-graduação e no momento da defesa do trabalho, pelas bancas examinadoras.

Os critérios de qualidade variam conforme as áreas de conhecimento, as instituições e até a banca examinadora em si mesma. Somente alguns critérios são de reconhecimento geral: tema interessante, clareza do discurso e procedimentos confiáveis. A partir desses critérios amplos, o pós-graduando vai construindo o seu discurso, colocando-o sistematicamente à prova através do seu orientador. Obras de metodologia científica são consultadas, geralmente, no decorrer da disciplina de "Metodologia...", quando essa faz parte das disciplinas obrigatórias, ou quando o orientador as indica.

Escalas de avaliação de pesquisas

Uma avaliação mais criteriosa de trabalhos científicos pode ser efetivada com a utilização de escalas de avaliação, que estabelecem itens básicos ou necessários a uma pesquisa científica, valorando-os. Como exemplo dessas escalas tem-se a "Escala de Greenville" (Witter, 1995b), desenvolvida em 1958, a "Escala de Suydam" (Witter, Silveira & Moraes, 1979), desenvolvida em 1968, a "Escala da AERA" (Ward, Hall & Schramm, 1975), desenvolvida em 1967 e modificada por Ward, Hall & Schramm, em 1975, e a escala desenvolvida por Ali, Young & Ali, apresentada em 1996. A finalidade dessas escalas é exposta por Witter, Silveira e Moraes (1979: 28), ao comentarem as escalas de Suydam e de Ward, Hall & Schramm:

"As referidas escalas servem para comparar o nível das pesquisas; para identificar pesquisas que, pela precariedade com que foram conduzidas, carecem de réplicas cuidadosas antes que seus dados possam ser úteis; para a avaliação de projetos de pesquisa; para autoavaliação de relatos antes de sua publicação e como forma de nortear o treino para o desenvolvimento de habilidade para análise crítica de pesquisas. Pode-se recorrer ainda a esses instrumentos na análise para a determinação do mérito ou demérito de todo um programa de pesquisa, e tomada de decisão quanto à continuidade da disponibilidade de verbas para o mesmo."

Ward, Hall & Schramm (1975) realizaram um estudo com a proposta de determinar a qualidade dos artigos relativos à pesquisa educacional publicados em periódicos da área, no ano de 1971. Os autores utilizaram os mesmos procedimentos de uma pesquisa realizada pelo *American Educational Research Association* (AERA), através do seu *Comitee on Evaluation of Research*, em 1967, que analisava os trabalhos publicados no ano de 1962, considerando a validade e a significância dos dados que proviam os educadores dos relatos de pesquisa. Na pesquisa da AERA foram encontradas pesquisas com falhas sérias, sendo considerado que apenas 7% das pesquisas eram dignas de publicação (conforme Ward, Hall & Schramm, 1975). Da mesma forma, no estudo de Ward, Hall & Schramm (1975), apenas 8% dos relatos (121 artigos como amostra em um

universo de 1486) foram considerados aceitáveis para publicação, 31 % aceitos após revisões menores, 34% aceitos após revisões maiores e 27% rejeitados. A avaliação foi realizada por professores doutores (atuantes na área de metodologia da pesquisa), com ampla experiência em pesquisa, significativa produção científica, vivência em corpo editorial e ampla experiência em orientações de teses.

A escala de Ward, Hall & Schramm analisa 33 características desejáveis como aspectos de qualidade na condução e relato da pesquisa. A escala da AERA continha 25 características desejáveis; a essa escala os três autores adicionaram mais 8 características. Para cada característica, como nas pesquisas relatadas, uma escala de cinco pontos foi usada, representando cinco níveis de qualidade: (1) Excelente (um modelo de boa prática); (2) Bom (poucos defeitos menores); (3) Medíocre (regular, nem bem, nem mau); (4) Pobre (alguns sérios defeitos) e (5) Completamente incompetente (um exemplo horrível) (Ward, Hall & Schramm, 1975: 116).

Os autores também verificaram relações entre as características dos juízes com suas avaliações. Os juízes membros da *American Psychological Association* (APA), os que haviam ministrado cursos de pesquisa educacional e os que haviam orientado teses produziram avaliações com escores significativamente baixos. Já os juízes que tinham experiência em pesquisa e número de artigos publicados, além da média, produziram avaliações com altos escores.

Ambos os grupos podem ser considerados ótimos pares para a avaliação de pesquisas, o que mais uma vez vem provar a necessidade de se estabelecer critérios absolutamente claros e delimitados para os sistemas de avaliação de pesquisas, afim de que se façam avaliações justas e similares.

A escala de Ward, Hall & Schramm (1975) divide-se em 33 características distribuídas em 7 categorias: categorias relativas ao título, ao problema, à revisão de literatura, aos procedimentos, à análise de dados, às conclusões, e à forma e estilo. Adotando-as, os autores acabam por nortear o discurso científico, ou seja, os relatos científicos somente serão considerados "de qualidade" se apresentarem todos esses passos.

Witter (1995a) incluiu mais 3 itens à escala de Ward, Hall & Schramm, os quais avaliam as referências bibliográficas apresentadas no trabalho científico, afim de se analisar as fontes de informação utilizadas pelo autor do trabalho: *As referências são apresentadas paradigmaticamente (seguem uma norma fixa e constante)*, *Temporalidade das referências (observar o princípio da temporalidade)* e *Tipologia do suporte documental (privilegia literatura cinzenta e periódicos)*.

Em trabalho de 1979, Witter, Silveira & Moraes relatam os procedimentos e resultados da comparação da escala de Ward, Hall & Schramm com a escala Suydam. A escala Suydam foi desenvolvida por Suydam nos anos sessenta, na *Pensilvania State University*, tendo tido amplo uso entre pesquisadores (Witter, Silveira & Moraes, 1979: 28), apresentando 9 itens que focalizam os vários aspectos dos relatos científicos. As autoras verificaram a equivalência entre as duas escalas tanto da parte teórica quanto da metodológica, sendo indicada a utilização de ambas as escalas como instrumentos paralelos como forma possível e eficiente de avaliação. A mesma pesquisa indica, paralelamente, a validação de uma escala pela outra.

Similarmente, Ali, Young & Ali (1996) apresentam, em outra escala, quase as mesmas categorias e critérios para determinar a qualidade de publicações e pesquisas em processos e decisões de promoções e nomeações em universidades norte-americanas. Os autores afirmam que normalmente as pesquisas dos candidatos são avaliadas a partir das características (internas) dos

trabalhos, como delineamento e estilo de escrita, e da reputação e status dos periódicos que os publicaram. Os autores apresentam em seu trabalho uma lista de itens ("*perceptions and variables*"), estruturada em uma escala de pontuações, que apontaria a qualidade de relatos de pesquisa publicados em periódicos, a qual poderia ser utilizada no processo de avaliação de promoções e nomeações. Consideram que a escala apresentada pode ser modificada conforme o interesse e as peculiaridades de cada instituição.

São nove os itens de qualidade (de artigos de periódicos propriamente) apresentados por Ali, Young & Ali (1996): originalidade da pesquisa (valendo até 25 pontos); método de pesquisa e testes estatísticos aplicados apropriadamente (15 pontos); estilo de escrita, interpretação e apresentação dos resultados (10 pontos); adequação das ilustrações (5 pontos); adequação das referências (5 pontos); autoria dos artigos (15 pontos); formato (10 pontos); natureza do trabalho (10 pontos); e tamanho (5 pontos). A pontuação é dividida em: "a" para *excelente*, "b" para *muito bom*, "c" para *bom*, "d" para *satisfatório* e "e" para *não satisfatório*.

A originalidade da pesquisa refere-se à abertura de novos campos de exploração, à replicação de uma utilizando um novo procedimento, à extensão e aprofundamento de pesquisas anteriores, e até à sumarização histórica que provê novos *insights* à disciplina. Note-se que este item é o que tem maior número de pontos (25% do total, enquanto na escala de AERA o item *significância do problema* atinge apenas de 3% da pontuação máxima), demonstrando sua importância em relação aos demais. Em relação ao segundo item (adequação do método...), os autores entendem que a questão está além da apropriação do método, devendo ser particularmente examinado o desenvolvimento de novos métodos e instrumentos que possam contribuir aos resultados de pesquisa. Em relação ao estilo de escrita, deve ser considerada a boa e clara apresentação dos fundamentos da pesquisa, até em relação à audiência. As ilustrações devem ser claras e apropriadas. As referências devem seguir os padrões do periódico, apresentando cerca de 10 a 22 referências (indicam ser mais o menos o padrão). Em relação à autoria, esta sendo única, deve receber a pontuação máxima, e sendo múltipla, receberá o primeiro autor um "b", no máximo, e assim por diante. Em relação ao formato da publicação, sua pontuação dependerá da complexidade da pesquisa e suas contribuições à disciplina: "a" para a discussão de problemas complexos, "b" para artigos de menor complexidade, "c" para revisões do estado-da-arte, "d" para análises repetitivas que não desenvolvem novas técnicas analíticas ou provêm novos dados para pesquisas futuras, e "e" para avaliações históricas. Em relação à natureza do trabalho, os autores apontam maior valor (10 pontos) para textos completos, médio para comunicações em eventos (5 pontos) e notas técnicas (3 pontos), e menor para comunicados curtos (1 ponto). O tamanho da pesquisa também é valorizado pelos autores, que indicam maior pontuação para trabalhos que contenham 5 mil ou mais palavras e menor para trabalhos de 3 mil ou menos palavras.

A escala de Ali, Young & Ali (1996) parece privilegiar aspectos mais externos ou formais do relato de pesquisa, enquanto que a escala AERA privilegia a exposição dos vários passos da pesquisa e suas adequações, desprezando questões relativas à forma, natureza e tamanho da pesquisa (na AERA estaria implícito que qualquer que fosse o relato - em qualquer tamanho e forma - ele deveria apresentar os mesmos procedimentos e passos. Ao mesmo tempo, a primeira escala valoriza imensamente a originalidade da pesquisa (representando 25% dos pontos), mas também muitos pontos discutíveis como a supervalorização da autoria única quando vem crescendo a valorização do trabalho em grupo, em colaboração em linha de pesquisa partilhada por muitos

pesquisadores.

Conclusões

Meltzoff (1997: 29) observa que a consistência do problema de pesquisa, o estabelecimento das hipóteses, do delineamento, da análise e das conclusões são os itens que o leitor espera de um bom trabalho. Esses são os itens básicos do que se intitula *pesquisa*. Não que os estudiosos não devam produzir outro tipo de trabalho, ou mesmo não deixar sua criatividade alçar vôo. Devem, sim, observar alguns critérios em suas produções, como expostos neste trabalho. Livros, artigos, notas técnicas, comunicações, teses, todos são produto da atividade Ciência. Como toda atividade algum controle é necessário, parâmetros são necessários embora seja também necessário um limite ao controle, para que não se crie mais um espaço autoritário e fascista - seria o fim da Ciência, que é, basicamente, uma atividade criadora; mas também é uma atividade específica, com uma linguagem específica. Especificidades essas que distinguem a Ciência de outras atividades e discursos.

O controle da atividade científica já está estruturado em agências, cursos superiores e corpos editoriais. Critérios claros e definidos de avaliação, a utilização das escalas demonstradas ou algo similar representam o controle do controle. É importante que a avaliação da produção científica tenha esclarecidos e públicos os seus critérios, a fim de que os estudiosos não tenham projetos de qualidade e importância social trocados por "politicagens" e preferências regionais ou institucionais. Portanto, toda a comunidade deve ser chamada a participar da elaboração desses critérios, ou reclamar a sua participação. Afinal, as boas idéias devem ser sempre bem-vindas.

Referências

- Ali, S. N.; Young, H. C. & Ali, N. M. (1996). Determining the quality of publications and research for tenure or promotion decisions. *Library Review*, 45(1), 39-53.
- Bruer, I. T. (1985). Methodological quality and citation frequency of the continuing medical education literature. *Journal of Documentation*, 41(3), 165-172.
- Bufrem, L. S. (1996). Linhas e tendências metodológicas nas dissertações do mestrado em ciências da informação do Instituto de Informação em Ciência e Tecnologia - Universidade Federal do Rio de Janeiro (1972-1995). Em: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 9, 27 out. a 01 novo de 1996, Curitiba. *Anais...* (Parte 4.4). Curitiba: UFPr: PUE.
- Campos, R. (1996, setembro, 22). A brincadeira de Sokal... *Folha de São Paulo*, n.p. (96214.)
- Castro, C. de M. (1986). *Ciência e universidade*. Rio de Janeiro: Zahar.
- _____. Há produção científica no Brasil? (1985). *Ciência e Cultura*, 37(7). Suplemento: Universidade Brasileira: organização e problemas.
- Castro, M. H. de. (1992). *Produção científica dos docentes da Escola Superior de Agricultura de Lavras: análise quantitativa*. Dissertação de mestrado não publicada, Pontifícia Universidade Católica de

Campinas, Campinas, E. São Paulo, Brasil.

Eco, U. (1989). *Como se faz uma tese*. São Paulo: Perspectiva.

Fazenda, I. C. A. (1991). Dificuldades comuns entre os que pesquisam educação. Em: FAZENDA, I. C. A. (Org.), *Metodologia da pesquisa educacional*. 2. ed. aum. (13-20). São Paulo: Cortez.

Foresti, N. A. B. (1990). Contribuição das revistas brasileiras de Biblioteconomia e Ciência da Informação enquanto fonte de referência para pesquisa. *Ciência da Informação*, 19(1), 53-71.

Forum on peer reviews and editors (1998). *Ciência e Cultura*, 50(1), 14-23.

Galembeck, F. (1990). Sem avaliações, sem progresso. *Ciência e Cultura*, 42(9), 627-628.

Guimarães, R. (1992). *Os mecanismos de avaliação em ciência e tecnologia no Brasil*. [S.l., s.n.]. (III-10. ZIP). Set. 1997.

Kostoff, R. N. (1996). Performance measures for government-sponsored research: overview and background. *Scientometrics*, 36(3), 281-292.

_____. (1997). *The handbook of research impact assessment*. 7. ed. Arlington: Office of Naval Research. DTIC Report Number ADA296021., 30/01/1998.

Lakatos, E. M. & Marconi, M. de A. (1995). *Metodologia do trabalho científico*. 4. ed. rev.e aum. São Paulo: Atlas.

Le Pair, e. (1995). Formal evaluation methods: their utility and limitations. *International Forum on Information and Documentation*, 20(4), 16-24.

Luma, S. V. de. (1991). O falso conflito entre tendências metodológicas. Em: FAZENDA, I. e. A. (Org.), *Metodologia da pesquisa educacional*. 2. ed. aum. (23-33). São Paulo: Cortez.

Macroberts, M. H. & Macroberts, B. R. (1996). Problems of citation analysis. *Scientometrics*, 36(3),435-444.

MeUo,1. M. e. de (1997). *Uma avaliação do sistema de revisão pelos pares ("peer review") no P ADCT*. S.n.t. (EPADCTOI.ZIP). (09/10/1997). 18 p.

Meltzoff,1. (1997). *Critical thinking about research: psychology and related fields*. Washington: American Psychological Association.

Meneghini, R. (1988). Indicadores alternativos de avaliação de desempenho científico: aplicação a uma instituição. *Ciência e Cultura*, 40(6), 548-558.

Meneghini, R. & Fonseca, L. (1990). Índices alternativos de avaliação da produção científica em bioquímica no Brasil. *Ciência e Cultura*, 42(9), 629-646.

Motta, D. F. da (1983). Validade da análise de citação como indicador de qualidade da produção científica. *Ciência da Informação*, 12(1),53-59.

PERFIL da pós-graduação. (1998). Brasília: CAPES. (Fev.1998.)

PERFIL da pós-Graduação: procedimentos básicos. (1999). Brasília: CAPES. (Jan. 1999).

Población, D. A.; Noronha, D. P. & Currás, E. (1996). Literatura cinzenta versus literatura branca: transição dos autores

- das comunicações dos eventos para produtores de artigos. *Ciência da informação*, 25(2), 228-242.
- Reformulação do sistema de avaliação da pós-graduação: o modelo a ser implantado na avaliação de 1998*. (1998). Brasília: CAPES. <http://www.capes.gov.br/perfil/reformula.htm>. (lan.1999).
- Spagnolo, F. & Sevilla, M. A. F. (1994). A situação atual da pós-graduação "lato sensu". *INFOCAPES*, 2(3), 7-20.
- Spinak, E. (1996). Los análisis cuantitativos de la literatura científica y su validez para juzgar la producción latinoamericana. *Boletín de la Oficina Sanitaria Panamericana*, 120(2), 139-146.
- Travassos, L. R. (1998). Editorial. *Ciência e Cultura*, 50(1), 7-8.
- Ward, A. W.; Hall, B. W. & Schramm, C. F. (1975). Evaluation of published educational research: a national survey. *American Educational Research Journal*, 12(2), 109-128.
- Witter, G. P. (1995a). *Escala de avaliação de artigo relatando pesquisa (ARRR, de Ward, Hall e Schramm)*. S.n.t. (Material distribuído durante aula da disciplina Análise da Produção Científica do curso de Mestrado em Biblioteconomia da PUC-Campinas).2.
- _____. (1995b). *Escalas de avaliação de pesquisas científicas*. S.n.t. (Anotações de aulas da disciplina Análise da Produção Científica do curso de Mestrado em Biblioteconomia da PUCCAMP, em 1995).
- _____. (1990). Pesquisa bibliográfica, pesquisa documental e busca da informação. *Estudos de Psicologia*, 1, 6-30.
- _____. (1997). Títulos de dissertações e teses em Biblioteconomia e Ciência da Informação (1972-1992). *Transinformação*, 9(3), 104-119.
- Witter, G. P.; Silveira, M. B. & Moraes, L. N. (1979). Comparação das escalas Suydam e ARRR para avaliação de pesquisa. Em: Witter, G. P. (Coord.), *Pesquisas educacionais*. (27-37). São Paulo: Símbolo.

Recebido em dezembro 1998.

Encaminhado para revisores em janeiro 1999.

Aprovado em janeiro 1999.