

Articles

Assessing Legal Research: Sense and Nonsense of Peer Review versus Bibliometrics and the Need for a European Approach

By Rob van Gestel & Jan Vranken*

Abstract

There is no uniform approach to assessing scholarly publications in legal research. Peer review is still the most accepted form, but the popularity of bibliometrics, such as the impact scores of journals and the citation scores of individual articles is increasing. However, there is no ranking of European law journals and none is likely to materialise any time soon. Comparing the UK, Belgium and the Netherlands, we will demonstrate why both peer review and the use of metrics have serious shortcomings. We believe it is necessary to think about such alternatives as more attention for methodological justification in legal research, more clarity from editorial boards about the quality criteria being used to approve or reject submissions, and more emphasis on standards for different forms of legal scholarship. Last but not least, we call for a Europe-wide debate on the pros and cons of different systems of research assessment, rather than let every country reinvent the wheel.

*Only a cynic knows the price of everything and the value of nothing*¹

A. Introduction

A fairly recent development in legal scholarship is the discussion on the assessment of the scientific quality of scholarly publications. In other sciences this debate took place decades ago. In his study on measuring scholarly research performance, Mark Pen observes how²,

* Rob van Gestel (r.a.j.vanGestel@uvt.nl) is professor of Theory and Methods of Regulation at Tilburg University. Jan Vranken (j.b.m.Vranken@uvt.nl) is professor of Methodology of Private Law at Tilburg University. The authors have established a *Research Group on Methodology of Law and Legal Research*, available at: www.tilburguniversity.nl/legalmethodology, in order to raise awareness and stimulate debate about the quality of law research. They wish to thank John Bell, Roger Brownsword, and Matthias Siems for their useful comments on a previous version of this article.

¹ Paraphrased from OSCAR WILDE, *LADY WINDERMERE'S FAN*, Act III (1892).

² See MARK PEN, *PRESTATIEMETING VAN WETENSCHAPPELINK ONDERZOEK. EEN EMPIRISCHE STUDIE NAAR DE DOORWERKING VAN UNIVERSITAIRE PRESTATIEMEETSISTEMEN* (2009). (Performance Management of Academic Research: An Empirical Study

since the 1970s, the scholarly tradition of restricting substantive assessment almost exclusively to peers has expanded to include quantitative metrics (bibliometrics). These include the ranking of journals and publishers, citations and other possible impact factors like the kind of publication (national or international) or the publication medium (article, book, book review, case note, etc.). The system usually works by assigning more credits to a publication with higher rankings or a higher number of citations.

Peer review is often preferred to evaluate the quality of legal publications because of different types of market failures that accompany a purely metrics-based system. For example, important academic success will sometimes only be recognized after long delay. Further, scientific innovation may occur in totally unexpected ways, as research answers other questions than those raised originally (the 'serendipity effect'). Because scientific discoveries are unpredictable, peer review instead of the law of the market is often believed to be the best way to assess if a certain publication represents a new and valuable contribution to the state of the art.

The assumption of those in favor of peer review that the most esteemed journals or publishers of the discipline have the highest threshold for acceptance, and consequently offer the best publications is not always correct in practice.³ Nevertheless, peer review is, for most legal scholars, the least controversial method of evaluation of the quality of scientific publications.⁴

The use of peer review and metrics is far less common than it is in other sciences in the European legal community. Until quite recently, even the notion of an explicit and independent assessment of the quality of legal research was unusual. While law journals have always had editorial boards responsible for accepting contributions, not all of these work with independent referees, and most do not make their assessment criteria explicit.⁵

Concerning the Impact of Academic Performance Assessment Systems. The dissertation is available full text and with an English summary, available at: <http://repository.tudelft.nl/> (last accessed: 8 March 2011).

³ See Dan Jerker B. Svantesson, *International Ranking of Law Journals: Can it Be Done and at What Costs?* 29 LEGAL STUDIES 681 (2009). The author even argues that truly ground-breaking work may be more likely to appear from marginal, dissident or unexpected sources, rather than from well-established and entrenched mainstream publication media.

⁴ See STEVEN WOODING & JOHN GRANT, *ASSESSING RESEARCH: THE RESEARCHERS VIEW* (2003).

⁵ This even applies to high-reputation journals, such as the *Modern Law Review*, *Law Quarterly Review*, *Oxford Journal of Legal Studies* and the *Cambridge Law Journal*. None of these journals publish which content-based assessment criteria are being used in the peer review process. Nevertheless, it is widely perceived that the reputation of the journal in which research is published affects how that research is assessed in the national research assessment exercise. See Kevin Campbell *et al*, 26 JOURNAL OF LAW AND SOCIETY 470, 349-350 (1999).

As a result, a harmonized and transparent system of research assessment of scholarly publications does not exist yet.⁶

For a variety of reasons (see under B), the pressure to establish more explicit criteria and procedures for the general assessment of the scientific quality of legal publications is likely to increase over the next few years. Must we expect a combination of peer review and metrics in legal research, much like in other social sciences? What are the generally acknowledged pros and cons of such systems in other disciplines, and what can the legal discipline learn from them when establishing its own assessment system for scholarly publications? Our point of departure is the emerging debate in the United Kingdom, Belgium and the Netherlands on the essential criteria for the assessment of legal publications, which are content-based and normally require some sort of assessment by peers. The virtues of peer review versus metrics in the assessment of scholarly European publications have been debated most intensively in the legal research communities of these three states, making them ideal case studies.

I. Delineation

The points of focus in this article will be the intrinsic quality of scholarly legal publications, its recognition through research assessment, peer review or through the application of metrics. Concentrating on the *intrinsic quality* of scholarly legal publications implies that the type,⁷ medium,⁸ and language of a publication are, in principle, irrelevant.⁹ That is why we leave aside in this article:

- If and how peer review and/or metrics can contribute to a reliable system of *ranking of journals or publishers*;
- What is the best way to assess the quality of legal *scholars*? Whether or not someone is a good scholar depends on more than just the content of his or her publications. Other quality indicators include editorships, awards, conference lectures, an international network, acquisition of research grants and subsidies,

⁶ By transparent we mean, among other things, that choice for a certain form of peer review by law journals is often unclear: Why do some prefer (double) blind reviews, while others favor that both the name of the authors and the reviewer are made public, and still others entrust the editorial board itself with the review of articles?

⁷ Type refers both to the kind of research, such as doctrinal, empirical or comparative research, and to the style in which the research is presented. Legal research can be formulated as an opinion, an article, an essay, a book, a case note etc.

⁸ Dissertations, monographs, articles, working papers, case notes, opinions, and such, in print or in an electronic version.

⁹ Obviously, language is an important factor when deciding for whom one is writing, and for the scope of the scholarly debate.

number of supervised PhD projects, and so on. In measuring these reputation aspects bibliometric indicators are probably indispensable, but as we will argue below the link between scholarly reputation and intrinsic quality of publications is not self-evident.

- Measuring the quality of *research programs* carried out by a group of researchers. When looking at publications from the perspective of research programs, not only the quality of the publications counts, but also the productivity, relevance, and vitality of the program.

A last preliminary remark is that we realize that assessments of the intrinsic quality of scholarly legal publications can serve different functions. Peer review for journals is, for instance, quite different to peer review of collective research exercises, books and book proposals. As for journals, the ultimate question is whether the submitted work is publishable or not. The editor or board must choose between various submissions, all of which are publishable in principle. In such cases editorial review may involve considering factors beyond the quality of the submitted article, such as its likely interest to readers of the journal, or its fit with the journal's subject area or publishing aims. We will return to these considerations in section F below, where we will suggest possible alternatives for peer review and metrics.

II. Order of the Argument

Section B contains a brief sketch of the reasons for the emergence of assessment of legal publications. Under C we will focus on typical content-based quality indicators applied by peers in the assessment of legal publications. We will then discuss the meaning and scope of three widely accepted criteria for the assessment of publications through research assessment exercises we identified in our three case studies, but which we believe are more generally applicable throughout Europe.¹⁰ These criteria are *originality*, *thoroughness* and *profundity*. We will show how relatively easy or difficult it is to apply these criteria. In section D, we explore the extent to which other criteria such as the societal impact or the cross-border nature of a publication might be just as relevant for assessment. Finally, section E analyzes the generally acknowledged pros and cons of content-based criteria and metrics.

¹⁰ We deliberately confine ourselves to the situation in Europe because a comparison with, for example, the situation in the United States (US) would not only require far more elaborate research, but would also complicate the discussion due to, among other things, the existence of student edited journals, the divide between the faculties of a few top law schools and the rest in terms of the distance between legal scholarship and legal practice, and the far greater dominance of law and economics and empirical legal research if compared to most European countries.

Our findings will be evaluated under F, where we emphasize a dilemma in the assessment of legal publications based on the identified disadvantages of both peer review and metrics. Can this dilemma be avoided when the legal discipline establishes its own system for assessing individual publications? We will suggest some alternatives in section G, inviting colleagues from other European countries to engage in a debate on the need for a harmonized assessment system for scholarly legal publications.

B. Where Does the Need to Assess Legal Publications Come From?

How is it possible that scholarly legal research has managed to stay out of the picture for so long in the debate on research assessment and performance indicators? And why is this now starting to change? Without being exhaustive, we will briefly mention some probable reasons.

First, at the national level, the importance of providing evidence to governments that their investment in research provides good value for money in terms of quality and impact has grown, especially in times of expenditure cuts.¹¹

Second, legal research was, and to some extent still is, closely linked to the nation state. Doctrinalists in particular often write about 'their' legal system: German private law, French administrative law, American constitutional law and so forth. Of course there is a long tradition of literature devoted to international law and comparative law, but the bulk of legal research was state-centric. Due to globalization, the rise of new 'law &...' disciplines, such as law and economics, and the increase of electronic journals, blogs, and other Internet platforms, legal research is spreading its wings like never before. In combination with the diversity of editorial requirements of journals and the various ways in which articles and books are being reviewed, these developments press the need for harmonized quality indicators for legal publications.¹²

Third, the popularity of national and international research assessment exercises also results from the increasing size of the academic community, where personal reputation is no longer sufficient to ascertain if someone is doing good work. Moreover, the sheer

¹¹ KEY PERSPECTIVES LTD., *A Comparative Review of Research Assessment Regimes in Five Countries and the Role of Libraries in the Research Assessment Process*, 8 (2009), available at: <http://www.oclc.org/research/publications/library/2009/2009-09.pdf> (last accessed: 3 January 2011).

¹² Margit Osterloh & Bruno S. Frey, *Research Governance in Academia: Are There Alternatives to Academic Rankings?* 3 CESIFO Working Paper Series No. 2797 4 (2009), available at: <http://ssrn.com/abstract=1460691> (last accessed: 3 January 2011).

number of publications¹³ makes it increasingly difficult and time-consuming to rely on peer review only.

Osterloh and Frey argue that the tendency to place a greater emphasis on rankings and impact scores is influenced by the adoption of ideas derived from the concept of 'new public management', according to which scholars must be monitored as regular employees. The underlying rationale is that scholars will react positively to pay-for-performance schemes as 'normal' workers do. This assumption however, remains unproven, though it is unlikely that legal research will be able to avoid a wider trend towards performance measurement in academic research.

Further, legal scholars must increasingly compete with those from other sciences for research funding provided by national and European research foundations. Legal scholarship has great difficulty explaining to fellow scientists from other fields which research methodology is normally followed, and cannot afford the lack of a transparent and generally accepted system of research assessment.

Finally, discussions on the assessment criteria for the quality of scholarly publications, especially of the suitability of bibliometric indicators, are not limited to legal scholarship. For instance, the European Research Foundation is currently trying to establish a *European Reference Index for the Humanities*, whose initial aims are to identify and increase the visibility of top-quality European humanities research published in academic journals in all European languages. In the humanities, claims for an exceptional position due to a deviating publication culture (e.g., the importance of books and non-English publications) have not been accepted as an excuse for making the quality of scholarly publications more transparent and internationally comparable.

C. Three Content-Based Criteria: Originality, Thoroughness, and Profundity

In the United Kingdom, the Research Assessment Exercise (RAE) was originally designed to maintain and develop the strength and international competitiveness of research in UK higher education institutions (HEIs), and to promote high quality research in these institutions to guarantee the largest proportion of government funding for the best research. There was strong endorsement for the 2008 RAE to be based on expert review by discipline-based panels considering submissions from HEIs. Explicit criteria were formulated for each subject area to enable the proper assessment of applied, practice-based and interdisciplinary research. For legal research this meant that the sub-panel for

¹³ For example, even as early as 2001 the *UK Research Assessment Exercise* assessed over 2,400 submissions and examined over 150,000 publications.

law considered each of the following quality profiles in direct relation to the general criteria of *originality*, *significance*, and *rigor*:¹⁴

- Quality that is world leading (4*): Research output that is, or is likely to become, a primary reference point in its sub-field.
- Quality that is internationally excellent but which nonetheless falls short of the highest standards of excellence (3*): Research output that is, or is likely to become, a major reference point that substantially advances knowledge and understanding in its sub-field.
- Quality that is recognized internationally (2*): Research output that is, or is likely to become, a reference point that advances knowledge and understanding of its (sub-) field.
- Quality that is recognized nationally (1*): Research output that makes, or is likely to make, a contribution to knowledge or understanding of its sub-field.
- *Unclassified*: Quality that falls below the standards of nationally recognized work or which does not meet the published definition of research.

In Belgium and the Netherlands, research assessments of publications also are still firmly based on peer review. Both the 2004 Flemish Report by the Verbeke Committee, the Model for the Integral Quality Assessment of Research in Law,¹⁵ and two Dutch reports, *Oordelen over Rechten* (Assessing Legal Research, Stolker Committee 2005)¹⁶ and *Naar Prestatie-Indicatoren voor Rechtswetenschappelijk Onderzoek* (Towards Performance Indicators for Legal Research, Smits Committee 2007)¹⁷ confirm this preference for a substantive assessment of publications. By and large they also agree¹⁸ that good scholarly legal publications should at least be *original*, *thorough*, and *profound* on a scale of 1 to 5.¹⁹ These criteria show great resemblance with the above-mentioned indicators used by the expert panels in the UK RAE 2008. Other possible qualitative indicators that were

¹⁴ In other words: the criteria of originality, thoroughness and profundity apply right across this range, weakly at the bottom, and paradigmatically at the top.

¹⁵ Vlaamse Interuniversitaire Raad (Flemish Inter-University Council, VLIR), *Model voor Integrale Kwaliteitsevaluatie van het Onderzoek in de Rechtswetenschappen* (2004). (The Verbeke Committee Report 2004). Available at: http://www.vlir.be/media/docs/Onderzoeksbeleid/notitie_KZR_22sept2004.pdf (last accessed: 3 January 2011). The Model is aimed at the assessment of individual scholars, not merely their publications. See under A.I for an explanation of the difference. We concentrate on the criteria to assess individual publications.

¹⁶ The Stolker Committee notes that indicators that are extensively used in other disciplines, such as a ranking of journals and citation analysis, are almost completely absent in (Dutch) legal scholarship.

¹⁷ Published by the Association of Universities in the Netherlands (VSNU), but no longer available through their website. On file with the authors.

¹⁸ *Stolker Committee Report*, 30 (2005); *Smits Committee Report*, 12 (2007). No longer available online. On file with the authors.

¹⁹ *Supra*, note 14.

mentioned in Belgium and the Netherlands are the *societal impact* and *cross-border* nature of publications (international or multidisciplinary). In the UK the 'international status' of publications is already embedded in the three criteria, while it is not in the other two cases. We will discuss this further under section E.

I. Originality

Originality may be juxtaposed with the reproduction and descriptive presentation of primary sources and existing literature. Original research shows that the author has made a personal contribution to the 'state of the art' of the research in a certain field. In other words, the research contributes something new to the scholarly debate on a subject. Originality could be connected to the research results, but this is not a requirement. Raising a new question could be original as well, as can choosing an innovative angle or research method, using new source material, or even interpreting existing sources so that they shed a different light on a familiar problem.²⁰ Originality is closely connected to theory building. This does not mean that scholarly publications must always contain a completely new theory. In fact, this is rarely the case. The more usual approach taken by scholarly legal research is to critique a doctrinal opinion, or to reveal inconsistencies in the literature.²¹ Another common motivation for scientific publications is charting a relatively new research subject, such as the present debate on the effects of neurobiology on law.²² All these publications can be scientifically original. Therefore, the term originality should be broadly interpreted, as the accumulation of knowledge is not a requirement. The restructuring or interpretation of existing material can be original as well. As always, there are of course degrees of originality.²³

²⁰ An example of the latter is revisiting jurisprudence 5 or 10 years later, to see whether the conclusions of an earlier investigation are still valid. Should this not be the case (wholly), the most interesting theoretical question becomes, 'Why not?'

²¹ See MARK TUSHNET, *WEAK COURTS, STRONG RIGHTS: JUDICIAL REVIEW AND SOCIAL WELFARE RIGHTS IN COMPARATIVE CONSTITUTIONAL LAW* (2008) for a clear example of the former. In this book, Tushnet argues that the US has few constitutional guarantees of social welfare rights such as income, housing, or healthcare. Legal scholars usually believe that the courts cannot possibly enforce such guarantees. By taking a comparative perspective Tushnet turns against this common belief and suggests that such rights can be judicially enforced, not by strengthening the power of the courts (stronger review), but by doing just the opposite, notably creating weaker forms of judicial review where the legislature and judiciary have more room to interact.

²² See, for example, the following special issue: 2 *INTERNATIONAL JOURNAL OF LAW IN CONTEXT* 217 (2006), and further on law and neuroscience.

²³ Hence the classification in the UK RAE and the five points scale in the Netherlands.

Although originality is a requirement of nearly every assessment protocol, both national and international,²⁴ the manner of interpretation of this criterion is not a matter of strict objectivity. Siems has shown that there are at least four different types of originality, depending on the kind of research and the internal or external perspective employed.²⁵ He pleads for a liberal attitude, in which one kind of legal research is not automatically assumed to be more original than the others. We concur and consider Siems' position to be important because legal scholarship as a field that is not free of trends; what is *en vogue* today can fade into obscurity tomorrow to make room for new styles of research.

For example, someone who conducts multidisciplinary or empirical research into the preventive effects of civil liability is most unlikely to unearth a fresh find through a purely dogmatic study of the relationship between unforeseen circumstances, and the adjustment or alteration of the positions of contracting parties. More often than not, the reverse will unfortunately be just as true. We can see this in the present debate on the unique nature of legal scholarship, where empiricists and multidisciplinarians cast doubt on the methods, and hence, on the scientific validity of doctrinalist research.²⁶ Doctrinalists are often tarred with the same brush. They are seen as deriving their arguments solely from authoritative sources, treating law as a coherent system in which decisions in individual cases are supposed to exceed arbitrariness simply because they must fit into the system. In reality however, doctrinalists come in different shapes and sizes. Some favor a purely internal perspective, others just as easily shift to an external (non-participant) perspective, still others rely heavily on authoritative sources and some consider facts and existing empirical data. Further areas of focus include the national legal system, a comparative law approach in research, and so forth. Therefore it is probably no wonder

²⁴ *The Standard Evaluation Protocol (SEP) for Public Research Organizations of VSNU, NWO, and KNAW* 9 (2003-2009), for example, describes quality as: 'international recognition and innovative potential'. See also the above mentioned UK RAE-criteria, and the international benchmark study conducted by Aldo Geuna & Ben R. Martin, *University Research Evaluation and Funding: An International Comparison*, 41 *MINERVA* 277 (2003). See also, *Expert Advisory Group for an RQF, The Research Quality Framework: Assessing the Quality and Impact of Research in Australia* 11 (2005), available at: http://www.dest.gov.au/NR/rdonlyres/E32ECC65-05C0-4041-A2B8-75ADEC69E159/4467/rqf_issuespaper.pdf (last accessed: 3 January 2011).

²⁵ Mathias M. Siems, *Legal Originality*, 28 *OXFORD JOURNAL OF LEGAL STUDIES* 174 (2008). He distinguishes between micro legal research, macro legal research, interdisciplinary research, and research on non-legal topics. Interdisciplinary research can in turn be subdivided into many forms. See also Bart van Klink & Sanne Taekema, *Dwarsverbanden. Interdisciplinair onderzoek in de rechtswetenschap*, 84 *NEDERLANDS JURISTENBLAD* 2563 (2009).

²⁶ See for example Deborah L. Rhode, *Legal Scholarship*, 115 *HARVARD LAW REVIEW* 1327, 1340 (2002), who writes: 'Yet too much conventional legal analysis is not done well. It exhaustively exhumes unimportant topics or replicates familiar arguments on important ones. Too little effort is made to connect law to life by assessing the real world consequences of analytic frameworks. Of course, to do so in systematic fashion would require significant time, money, and expertise, which is precisely what most authors of doctrinal works are happy to avoid. The result is that, on many key legal issues, we are glutted with theory and starved for facts.' See also Gerrit A. de Geest, *Hoe maken we van de rechtswetenschap een volwaardige wetenschap?* (How to transform the legal discipline into a real science?) 79 *NEDERLANDS JURISTENBLAD* 58 (2004).

that doctrinalists refute the criticism, and wonder whether a large part of non-normative research deserves to be called *legal* scholarship at all.²⁷

The vehemence in this debate is probably connected to the fact that lawyers who primarily conduct doctrinal research now feel that their territory is being subsumed as they are seen as old-fashioned and hence, by definition, unoriginal.²⁸ Doctrinalists sometimes seem to forget that experience shows that those who desire change normally need to kick up the dust to get their message across.

II. *Thoroughness and Profundity*

Profundity is associated with visible competence and methodological rigor.²⁹ Pertinent questions address if the researcher overlooked relevant sources leading to an incomplete or distorted answer to the research question, and if the selected method (e.g., a study of case law or legislation) can suitably answer the research question (e.g., is the law effective in practice)?³⁰ Some are inclined to regard profundity and thoroughness as synonyms.³¹ We believe in a real difference between the two criteria. *Thoroughness* describes the extent to which authors should try to find underlying explanations or theories for developments in law (asking 'why' questions instead of 'how' questions).³² *Profundity* describes the extent to which the publication should provide a comprehensive answer to the research question through reliance on relevant sources. The foremost requirement for a research question is to point out what the investigation aims to contribute to the existing

²⁷ See Douglas W. Vick, *Interdisciplinarity and the Discipline of Law*, 31 OXFORD JOURNAL OF LEGAL STUDIES 163, 164 (2004), who claims that: 'Many interdisciplinarians perceive doctrinalists to be intellectually rigid, inflexible, and inward-looking; many doctrinalists regard interdisciplinary research as amateurish dabbling with theories and methods the researchers do not fully understand.' Some scholars argue that the true nature of legal research lies in the normative character, which implies that legal scholarship is all about how the law should read. For a clear example, see Jan M. Smits, *Redefining Normative Legal Science: Towards an Argumentative Discipline*, in METHODS OF HUMAN RIGHTS RESEARCH 45 (Fons Coomans ed., 2009). Others do not go so far but do criticize much of the interdisciplinary legal scholarship. See Neil Duxbury, *A Century of Legal studies*, in THE OXFORD HANDBOOK OF LEGAL STUDIES 950, 957 (Peter Cane & Mark Tushnet eds., 2003) referring to 'cross-disciplinary chutzpah.'

²⁸ An extensive description of doctrinal legal research can be found in MICHAEL SALTER AND JULIE MASON, *WRITING LAW DISSERTATIONS* (2007). See also, FIONA COWNIE, *LEGAL ACADEMICS: CULTURE AND IDENTITIES* (2004).

²⁹ See Pen (note 2), 3.

³⁰ For instance, when the research aims at increasing the practical functionality of (a part of) the legal system (making it more effective and efficient), but when its conclusions are based on nothing more than dogmatic interpretation of sources, making it impossible to answer the research question.

³¹ *Smits Committee Report 20, 27* (2007), *supra*, note 18. The book that the chairman of the committee, Jan Smits, later wrote, JAN M. SMITS, *OMSTREDEN RECHTSWETENSCHAP* (Contentious Legal Scholarship) (2009), shows how little he expects from methodological requirements.

³² The Latin root of profundity means depth.

body of knowledge. In this sense profundity is linked to the quest for originality, but it is also its basis.³³ It has its own distinctive rules. Profundity should not be confused with an obsession for details or expansiveness, such as with overly long or annotated publications. A paper of fewer than 5,000 words can have greater scientific relevance than an article exceeding 15,000 words on the same theme because of an original angle or way of reasoning.³⁴ This makes it rather fatuous to set minimum requirements for the number of words in papers, books, annotations, etc.³⁵ before they can be considered scholarly publications. A lengthy publication could just as easily be the product of a prolix writer.

Another question is whether profundity brings to light a difference in publication culture. Many articles in especially American student-edited law journals have countless footnotes, and are written in a style that seems tedious to many European scholars. They feel one often has to trudge through reams of sources to follow the line of reasoning. Yet, mostly, the answer to the research question in those journal articles can be found in the abstract as well as in the introduction and then again in the conclusion. Here, thoroughness might stand in the way of persuasiveness and readability, which is also important to scholarly legal research.

In short, profundity can both contribute to properly answering the research question and diminish creativity and originality. Another complicating factor for writers is the great variety in available forms of publication in the legal field. Is there more room for shorter opinions, essays or other pieces of outspokenness, so long as they contain at least some original thoughts? When do these publications fall short of true scholarly significance? What is the difference between 'armchair science',³⁶ fireside chats and idle musings?

D. Multidisciplinarity, Internationalisation and Societal Relevance as Independent Indicators?

For a substantive assessment of legal publications the type of legal research is not decisive of its quality. That is why we oppose those who implicitly or explicitly place a higher value

³³ Verbeke Committee Report (2004), *supra*, note 15, cited in the Stolker Committee Report 30 (2005), *supra*, note 18

³⁴ A good example is the renowned article by Garrett Hardin, *The Tragedy of the Commons*, 162 *SCIENCE* 1243, (1968), which comprises a mere 5 pages and 21 footnotes. Regardless, it is one of the most important articles in the field of (environmental) law & economics. Hardin's article serves as inspiration for, among others, the Nobel Prize-winning research by ELINOR OSTROM, *GOVERNING THE COMMONS: THE EVOLUTION OF INSTITUTIONS FOR COLLECTIVE ACTION* (1990).

³⁵ *Stolker Committee Report* 31 (2005), *supra*, note 18, mentioned 4,000 words, but only as a general and flexible guideline.

³⁶ Inspired by the former British popular science journal of the same name.

on certain types of research – especially international, multi- or interdisciplinary and empirical – over national and/or doctrinal research. All research, however, must live up to the same quality standards.

Further, we acknowledge that satisfying the three content-based quality criteria we identified in section C can be either more or less problematic for some types of research. For example, the originality requirement can be fulfilled differently. Doctrinal research aiming at interpreting recent case law or a legal provision has a harder time fulfilling the standard of originality than research into, as Siems calls it, macro-level questions.³⁷

I. Multidisciplinarity

Even more complicated is multidisciplinary research, which is often extolled because the borders between different fields inspire new insights much more easily than monodisciplinary research does. Yet, law is not unique in this regard. There are no scientific or scholarly fields where it does not or should not occur. It is true that the innovative strength of multidisciplinary work is often obvious, but this type of research has a harder time living up to the standards of thoroughness and profundity. Thorough and profound multidisciplinary and empirical research generally requires more methodological expertise and technical skills than its doctrinal or comparative legal counterparts. Law as a discipline in Europe is attempting to incorporate this methodological knowledge and research skills. While we endorse these efforts enthusiastically, there is also much that must be learnt with an open mind.³⁸

II. Internationalization

Internationalization offers the ‘existing overlap’ argument to refrain from the use of research type as an indicator of quality. Internationalization and Europeanization are cardinal issues nowadays, but there is no reason to assign them a fundamentally higher status than national research.³⁹ Research in the fields of international or European law usually only distinguishes itself from national legal-dogmatic research through its sources and, when it is presented abroad, through its forum and language, rather than through the methodology, approaches or reasoning applied. De Witte recently confirmed this, stating

³⁷ See Siems (note 25).

³⁸ The quality of research of this kind is the subject of much concern in the US as well. Compare among others, Richard Posner, *Legal Scholarship Today*, 115 HARVARD LAW REVIEW 1314, 1314-1326, 1323-1326 (2002), and the harsh criticisms leveled throughout their paper by Lee Epstein & Gary King, *The Rules of Inference*, 69 CHICAGO LAW REVIEW 1 (2002).

³⁹ Amongst others, the *Smits Committee Report* 32 (2007), *supra*, note 18.

that there is no 'relatively unified, cross national community' of scientists involving itself with European law, despite what conventional wisdom often holds.⁴⁰ The reality remains that even though comparable European legal questions are considered, the field is far from unified due to language barriers and differences in the training and cultural backgrounds of individual researchers operating within national legal systems. This same point was advanced in the UK, where the RAE linked 'significance' to 'international quality'.⁴¹ The Society of Legal Scholars emphasized that the panel responsible for implementing general assessment criteria in the legal field does not use the term 'international quality' in the geographic sense, as legal research in most cases is still closely tied to specific jurisdictions, which, even when broadly interpreted (as in the common law), are far from global. International quality therefore needs to be understood as being a 'primary reference point in its field'.⁴² This is a content-based quality factor for the assessment of publications, but one that raises new questions— when and how do we decide if a publication is a primary reference point for an entire legal field, and is this always evident right off the bat?

Profundity, originality, and thoroughness of international or European publications are not intrinsically superior in quality— research in these areas of law is either good, mediocre or even bad, as with all other research. Quality depends exclusively on the content, and not on the publication style or the medium. Well-known legal writers have managed to draw attention to groundbreaking research through a simple essay, without a single footnote. A case in point is Lon Fuller's article, 'The Case of the Speluncean Explorers,' published in the Harvard Law Review. Generations of American law students grew up with it, even though the essay violates just about every conceivable rule on methodological justification.⁴³ Nor does quality depend on the language of publication, since history tells us that high-profile work in various legal disciplines has been translated into English, often long after the initial publication date.

III. Societal relevance

As far as the societal relevance of legal publications is concerned, it cannot serve as an independent indicator of the quality of legal publications. As has been pointed out, this

⁴⁰ Bruno de Witte, *European Union Law: A Unified 'Academic Discipline?',* 34 EUI WORKING PAPERS RSCAS (2008), available at: <http://cadmus.eui.eu/dspace/handle/1814/10028> (last accessed: 3 January 2011).

⁴¹ For more extensive information, see Ron Johnston, *On Structuring Subjective Judgements: Originality, Significance and Rigour,* 62 HIGHER EDUCATION QUARTERLY 120, 120-147, 132-133 (2008).

⁴² *Society of Legal Scholars,* AHRB JOURNAL REFERENCE LIST (2005), available at: http://www.legalscholars.ac.uk/pubdocs/05/joint_letter_ahrb.pdf (last accessed: 3 January 2011).

⁴³ A complete reprint can be found online: Lon L. Fuller, *The Case of the Speluncean Explorers,* 62 HARVARD LAW REVIEW 4 (1949), available at: <http://www.nullapoena.de/stud/explorers.html> (last accessed: 8 March 2011).

factor does not belong to the academic standards of (legal) research. We are aware that perhaps there is no benefit from legal publications which completely divorce themselves from legal practice or societal needs and deteriorate into ivory-tower discussions between academics 'playing air guitar at a concert with no audience',⁴⁴ but one never knows whether, how and even when good or brilliant scholarly publications can become fruitful in society. Societal relevance can enhance the development of scholarly research, but it cannot be used as an indicator of its quality. In that respect, the two are not communicating vessels. Therefore, the logic behind the petition to the British Prime Minister to allocate funds for academic research solely on the basis of academic excellence and not on the basis of 'impact' or the judgments of users comes to light.⁴⁵

E. Are Bibliometrics More Useful for Assessing Scholarly Publications?

I. Possible Disadvantages and Negative Side Effects of Peer Review

A system based on substantive assessment appears to run into a number of practical obstacles,⁴⁶ especially where large research assessments are concerned with many scholars involved, as in nation-wide or institution-wide RAEs. Not only does reading all the key publications take an overwhelming amount of time, it is also quite problematic to find enough highly qualified peer reviewers who are willing to take on this job, which is often either voluntary work or modestly remunerated. National or European research foundations encounter the same problem. Often, hundreds of applications are submitted for only a few grants, which makes assigning all of them to independent reviewers all but impossible.⁴⁷ As a consequence of this, preselections are unavoidable, but how these take place is not really transparent to the outside world.

⁴⁴ The latter reproach is aimed at American legal scholars in, for example, Pierre Schlag, *Spam Jurisprudence, Air Law, and the Rank Anxiety of Nothing Happening*, 97 *GEORGETOWN LAW JOURNAL*, 803, Please insert the specific page no. of this citation, if available (2009). Even in the US, where 'law &...' research at top law schools now enjoys a higher status than doctrinal research, fervent supporters of multidisciplinary such as Richard Posner argue for a revitalization of doctrinal research. See RICHARD A. POSNER, *HOW JUDGES THINK* 211 (2008). More on the European context can be found in, for example, Anthony Arnull, *The Americanization of EU Law Scholarship* in *CONTINUITY AND CHANGE IN EU LAW* 415 (Anthony Arnull, Piet Eeckhout & Takis Tridimas eds., 2008).

⁴⁵ See E-Petitions, THE OFFICIAL SITE OF THE PRIME MINISTER'S OFFICE, available at: <http://petitions.number10.gov.uk/REFandimpact> (last accessed: 3 January 2011). The first signs are that the UK government might actually change the societal impact agenda proposed for the Research Excellence Framework. See David Willetts, *David Willetts delivers first keynote speech as Minister for Universities and Science*, DEPARTMENT FOR BUSINESS INNOVATION & SKILLS, available at: <http://www.bis.gov.uk/news/speeches/david-willetts-keynote-speech> (last accessed: 3 January 2011).

⁴⁶ For a good overview see, ANN C. WELLER, *EDITORIAL PEER REVIEW: ITS STRENGTHS AND WEAKNESSES* (2001).

⁴⁷ Research by Peter van den Besselaar & Loet Leydesdorff, *Past Performance, Peer Review, and Project Selection: A Case Study in the Social and Behavioral Sciences*, 18 *RESEARCH EVALUATION* 273, 288 (2009), showed that grants

Another problem regarding substantive assessment is that scholars can, in principle, be expected to be open, honest and free of prejudice when reviewing the work of their peers. In practice however, competing opinions, personal preferences, and conflicts of interest are not always avoidable. This is especially true when the scholarly forum consists of oft-competing colleagues from the same country, or when the field has only a limited number of international specialists.⁴⁸ In those cases one can also find 'peer review cartels' that are more focused on staying friends with other established colleagues and on functioning as 'gatekeepers' of mainstream theories than on stimulating innovative scholarship⁴⁹ by offering opportunities to young and unknown researchers and 'mavericks' who attempt to criticize established scholars in their field.

II. The Rise of Bibliometric Indicators

These two circumstances, together with our observations in sections B, C, and D have contributed to the growing acceptance of bibliometric indicators in legal scholarship.⁵⁰ As far as is relevant to the assessment of individual publications, the two top indicators are: publication in high ranked journals or by high ranked publishers, and citation indexes. In the UK, the 2008 RAE was still strongly based on expert review,⁵¹ but with respect to the use of citation information, a report from September 2009 notes that '[b]ibliometrics are not sufficiently robust at this stage to be used formulaically or to replace expert review in the REF (Research Excellence Framework, successor to the RAE). However, there is a considerable scope for citation information to be used to inform expert review.' But the balance between peer review and the use of metrics varies across the fields of research. In 'areas where publication in journals is the main method of scholarly communication,

for social and behavioral sciences do not end up with the best researchers in terms of publications in journals from the *Web of Science* and their citations measured over the recent period.

⁴⁸ These disadvantages of peer review are thoroughly documented. See, for example, MICHELE LAMONT, *HOW PROFESSORS THINK: INSIDE THE CURIOUS WORLD OF ACADEMIC JUDGMENT* (2009). We also refer to the results of the ISI Peer Review Survey, which can be found online: *Peer Review Survey 2009: Preliminary Findings*, SENSE ABOUT SCIENCE, available at: <http://www.senseaboutscience.org.uk/index.php/site/project/395> (last visited 3 January 2011); PARLIAMENTARY OFFICE OF SCIENCE AND TECHNOLOGY, PEER REVIEW, Postnote no. 182, 1-4 (2002), available at: <http://www.parliament.uk/post/pn182.pdf> (last accessed: 3 January 2011); and Johnston, (note 41), 120-147.

⁴⁹ J. Scott Armstrong, *Peer Review of Journals: Evidence on Quality Control, Fairness and Innovation*, 3 *SCIENCE AND ENGINEERING ETHICS* 63, 15 (1997). See also James Bradley, *Pernicious Publication Practices*, 18 *BULLETIN OF THE PSYCHONOMIC SOCIETY* 31, 34 (1981).

⁵⁰ More background can be found in PEN (note 2), 1-8 among others.

⁵¹ This is not without debate, however. See, for example, Stephen Bailey, *Reform of Higher Education Research Assessment and Funding : Response by the Society of Legal Scholars*, THE SOCIETY OF LEGAL SCHOLARS, University of Nottingham School of Law, available at: <http://www.legalscholars.ac.uk/pubdocs/responses/sls-metrics-2006.pdf> (last accessed: 8 March 2011).

bibliometrics are more representative of the research undertaken.⁵² In legal research, books are still highly important, although perhaps less so than a few decades ago when most law journals were (even) more nationally oriented.

In Australia, the *Research Quality Framework* was replaced in 2008 by the *Excellence in Research for Australia* (ERA) framework, which also combines metrics and expert review committees across eight disciplinary areas to ensure that researchers are not treated unfairly in fields where bibliometric data are not yet readily available, (e.g. legal research, arts and humanities).⁵³ In Ireland, where a nation-wide research assessment system does not yet exist, individual institutions compare their research performance with that of others through internal and external audits. Quantitative performance indicators, such as the ranking of the journal that carries an individual article, the impact factor of the journal itself or of the publication in the journal, the number of times the publication is being cited or downloaded (in case of electronic publications), are gaining ground in the process of assessing scientific publications in Ireland, as is shown by a 2009 international benchmark study.⁵⁴

In this respect, Denmark offers an interesting view. Work has begun there on the development of a system based on publication counts, which in turn is modeled on Norwegian bibliometrics.⁵⁵ The *Danish Council for Research Policy* has developed a framework that counts the number of peer-reviewed articles and books, journal rankings and citations. Nevertheless there are complaints, especially from fields including history and law that also depend on localized publishing practices, and for that reason often nominate Danish publishers and journals in the highest ranking.⁵⁶ Moreover (and this is

⁵² Report on the pilot exercise to develop bibliometric indicators for the REF (Executive Summary, 2009), available at: http://hefce.ac.uk/pubs/hefce/2009/09_39 (last accessed: 3 January 2011).

⁵³ See *The Excellence in Research for Australia (ERA) Initiative*. Available at: <http://www.arc.gov.au/era/default.htm> (last accessed: 3 January 2011). The ERA scheme also requires as one of its components international journal ranking. Svantesson (note 3), 680 is very critical about the linkage between assessing research quality and ranking.

⁵⁴ Forfás/HEA, *Research strengths in Ireland: A bibliometric study of the public research base* (2009). Available at: http://www.forfas.ie/media/forfas091209_bibliometric_study.pdf (last accessed: 3 January 2011). The report tracked publication and citation levels for Ireland's higher education institutions in the ten-year period to 2007. The analysis was based on a global database of 10,000 of the most highly cited journals in the world. All seven universities were included. During that period the number of peer-reviewed papers published per year more than doubled to approximately 5,000. The study also concluded that the overall citation impact of Irish research is improving.

⁵⁵ Hanne Foss Hansen, *Research Evaluation: Methods, Practice and Experience*, 52-60 DANISH AGENCY FOR SCIENCE, TECHNOLOGY AND INNOVATION (2009).

⁵⁶ See Key Perspectives Ltd. (note 11), 37. Svantesson (note 3), 690, on the other hand, points out, as one of the specific negative consequences of international journal ranking, that smaller jurisdictions will hardly be able to have national journals classed as top ranked journals, and that this will encourage scholars to publish in foreign

typical of most metrics-based systems), the Danish system does not solve the problem of direct causality between the quality of an individual publication and the credits it receives for being published in a certain journal or for the number of times it is cited (see section F for further discussion).

In Belgium and the Netherlands, the most effort was put into establishing a ranking of journals and publishers. Citation indexes are still completely absent in law and are unlikely to develop soon. In Belgium, the Verbeke Committee proposed ranking law journals and books according to their general or average scholarly content, assigning them an A, B, or C status, but this plan failed completely.⁵⁷ When the list was published, the ranking caused such controversy that the committee had to be dissolved.⁵⁸ No national assessment of legal research has since been undertaken in Belgium.

In the Netherlands, the Stolker Committee advocated a national system of classification that distinguishes between scholarly and professional publications, and between popular papers and reports. For certain publication types that are typical to legal scholarship, such as loose-leaf collections, case notes, thematic collections, chronicles, and country reports, the system provides criteria to ascertain whether these can be regarded as scientific.⁵⁹ The committee's successor, the Smits Committee, while acknowledging that the opinions of peers are personal, felt that it should be possible to develop indicators to ensure objectivity. It also assigned a central role to the ranking of journals, but it did not attempt to develop a ranking system, leaving this onerous and thankless duty of making classification decisions to a new body, the Du Perron Committee. So far, this last group has not reported its results, but believes (according to its chairman), that very few Dutch law journals fulfill all criteria and thus qualify as publishing papers that meet scholarly requirements. This may well be the main reason why the final report is still waiting to be published. The Belgium experience may have instilled caution in the committee.

III. Metrics and Quality of Individual Publications

In the previous section we mentioned the problem of causality between the credits a publication receives according to a metrics-based assessment system and the quality of the content of that same publication. As the Smits Committee has pointed out, the ranking of a

journals whose readers are not interested in domestic issues of the author. Moreover, international ranking will favor certain disciplines over others. For example, when many journals of one specialist area are highly ranked, research in that area is automatically higher ranked than in other areas with less top-tier journals.

⁵⁷ See *Verbeke Committee Report* (2004), *supra*, note 15.

⁵⁸ Serge Gutwirth, *Evaluatie rechtswetenschappelijk onderzoek*, 168 *Nieuw Juridisch Weekblad* 674, 677 (2007).

⁵⁹ *Stolker Committee Report*, 26-36 (2005), *supra*, note 18.

journal does indicate the average quality of the publications appearing in that journal, but says little about the scholarly quality of an individual contribution. To a certain extent it also transfers the problem from the quality of the publication itself to that of the journal. Why is a journal highly-ranked? Is it because many scholars read it, because it has, or is reputed to have a good peer review system, or because the articles in the journal are cited more often than similar publications in other journals on average? In other words who, in a metrics-based approach, determines what quality means and how?

As Pen's dissertation⁶⁰ shows, in fields where journals and publishers are ranked, quality judgments are largely based on the ranking. Scores are no longer separately assigned according to the substantive quality of publications. The quality of the individual publications is simply assumed. This seems logical, but is in fact an oversimplification with significant disadvantages, as outlined below.

Although citation analysis and other impact metrics are yet to be implemented in legal research in the UK, the Netherlands and Belgium, some law schools have begun to use a system which classifies journals according to their being national or international, refereed or non-refereed and so on. These faculties⁶¹ assign extra credits to individual publications on the basis of their international character (which sometimes means no more than that they are published in English), or that other than doctrinal methods are being applied (multidisciplinarity), or just because a respected house publishes them. These publications sometimes receive twice as many credits as articles without those features do. Furthermore, almost all law schools require their researchers to produce a minimum number of publications over a certain period of time, which in itself of course says nothing about the quality of those individual publications.

IV. Disadvantages and Negative Side Effects of Bibliometric Indicators

Quantitative standards, expressed in numbers, have the apparent advantage that they seem more unequivocal, objective and verifiable than substantive assessments that give a verbal description of quality. As mentioned in section A, appearances can be deceiving. How can originality of insight be quantitatively measured or expressed in numbers? How can one be sure that an article is not only cited for the sake of completeness, but because other writers are really influenced by it?⁶² And what would the right period for

⁶⁰ See PEN (note 2).

⁶¹ In the Netherlands, for example, the faculties of Tilburg University, Rotterdam University, and Amsterdam University.

⁶² Compare Abraham de Swaan, *Oordeel en vooroordeel; over verantwoorde beoordeling van onderzoek (Judgement and Prejudice; Judgement and prejudice; On Responsible Research Assessment)* in MACHT EN VERANTWOORDELIJKHEID: ESSAYS VOOR KEES SCHUYT 267, 271 (Jan Willem Duyvendak *et al.*, eds., 2007).

measurement be?⁶³ How can a publication by a beginning scholar be compared in a fair manner with one by an established scholar in terms of impact? How should contributions in books be compared to articles in (international) journals? Which international journals should be considered and which ignored— should English-language journals be given centre stage, or are journals in other languages just as important? These are all pertinent questions, leading to the necessity for formulation of many more similar ones.

Experience from other fields tells us that assessment based on metrics has numerous unwanted side effects in practice. In the case of Economics, competition with other scientists on a global scale has caused a knowledge paradox: a large part of the publications in specialized journals is too divergent from the needs of the real-world economy.⁶⁴ Comparable problems exist in the United States where legal research and education at top law schools are concerned. Some would say that much research conducted there is completely divorced from what legal practice needs. Furthermore, a rift between research and education could develop.⁶⁵

Using bibliometric indicators is also known to trigger strategic behavior, dissolving even an indirect relationship between the quality of the journal and the individual articles published in it. Examples include 'duplicating research output, avoiding risky research which would not immediately lead to publication, or dividing research results across multiple publications' (known as salami-style publishing).⁶⁶ Osterloh and Frey cite a phenomenon in Australia in the mid-1990s, where the remuneration of scientists and the funding of universities were linked to the number of publications in cited journals. The number of publications increased dramatically, but the quality measured by the number of citations dropped correspondingly.⁶⁷

The most basic risk of quantitative quality assessment is not strategic. Neither is it a game of cat and mouse between scientists and reviewers.⁶⁸ A much greater risk is a slow but

⁶³ The Social Science Citation Index (ISI) normally uses a period of one or two years. See the KNAW report, *Judging research on Its Merits* 18 (2005). Available at <http://www.knaw.nl/publicaties/pdf/20051029.pdf>. (last accessed: 3 January 2011). This document points out that where the humanities and social sciences are concerned, this period is usually too short to properly judge the impact of a publication.

⁶⁴ See KNAW, *supra*, note 63, at 13, 14.

⁶⁵ Harry T. Edwards, *The Growing Disjunction between Legal Education and the Legal Profession*, 91 MICHIGAN LAW REVIEW 34 (1992) caused quite a stir in the US.

⁶⁶ See Pen (note 2), IX, 5-7.

⁶⁷ Margit Osterloh & Bruno S. Frey, *Evaluitis: de ziekte van de 'gecontroleerde' wetenschap*, (Evaluitis – The Disease of 'planned' science) in IF YOU'RE SO SMART, WHY AREN'T YOU RICH? 235, 237 (Chris Lorenz ed., 2008).

⁶⁸ Compare the interview in NRC Magazine with the Dutch Spinoza Prize winners, who gave an acute description of how they manipulated the NWO grant system, for example by applying for grants for research proposals which

steady undermining of the freedom of scientists to determine and execute their own research agenda, which would result in drab uniformity instead of stimulating diversity and drive an even greater wedge between research and education. The UK Higher Education Policy Institute has also pointed this out, adding that if the RAE is seen as too bureaucratic and costly, the answer would not be a stronger emphasis on output control and competition (through the inclusion of a bibliometric research component in the RAE, which would result in even higher transaction costs), but rather, an increased reliance on the self-correcting ability of law schools and a proportional external quality control.⁶⁹

The preceding argument is balanced by Pen's empirical study⁷⁰ involving four Dutch faculties, where research funding was divided according to a performance measurement system. This study concludes that many of the negative effects mentioned in literature were not as pronounced as expected. This was achieved through two mechanisms: cushioning and compensation. Cushioning includes making exceptions, creating a time gap between performance and reward, and allowing fairness arrangements. Compensation consists of only partly dividing funding according to performance measurement and using other instruments such as start-up grants, incentives, career management, policy funding or other earmarked funds or reserves. Pen argues that metrics should be constantly debated, more than they are now, because all measurement systems unavoidably become less efficacious over time; researchers adjust and the system loses its stimulating and discriminating effect.

This prompts the following question: what problems are bibliometric research indicators really meant to solve? The relatively comforting results of Pen's study are no longer primarily concerned with furthering the scholarly quality of individual publications. The purpose seems to have shifted towards creating an instrument for oversight, management, and policy, which is just as ineffective in guaranteeing a lasting high quality of scholarly publications as substantive assessment by peers. Implementing both systems cumulatively would only add to the burden on the time and efforts of researchers to justify their work, leaving less time for research and education.⁷¹

had already been completed - which allowed them to give very precise descriptions of the research method and expected outcomes, and using the resulting money to finance new research. NRC Wetenschap June 4 2009.

⁶⁹ Tom Sastry & Bahram Bekhradnia, *Using metrics to allocate research funds: A short evaluation of alternatives to the Research*, 18 (2006), available at: <http://www.hepi.ac.uk/files/23RAEandmetricsfullreport.pdf> (last accessed: 3 January 2011).

⁷⁰ See Pen (note 2), 165-168.

⁷¹ According to the KNAW Report, *Kwaliteitszorg in de wetenschap* 6 (2008), existing quality assurance systems are seen as too much of a burden.

F. Can this Dilemma be Avoided?

Partly taking our cue from experiences in other scientific fields, we conclude that both substantive assessment of scholarly legal publications and employing bibliometric indicators can have significant disadvantages and adverse effects. Peer review is hindered by the obscurity of the assessment criteria, subjective beliefs and biases of reviewers, while metrics say little about the quality of an individual publication from a substantive perspective. A cumulative implementation of peer review and metrics will perhaps swell bureaucracy to an unacceptable level. How can this dilemma be avoided?

Legal scholarship has the advantage of a late start. As elaborated in section E, most Western European countries do not have a ranking of law journals and citation indexes yet. Even external peer review is still far from common, and the selection of reviewers, the criteria to be applied, and the accountability of both the reviewers and the editorial board for the assessment of publications generally remain unclear.⁷² To a certain extent this backlog can be seen as an advantage. It means that the constructions of scholarship review mechanisms can be done from the foundations, without uphill battles to change existing viewpoints. However, since some law schools are already working to implement metrics into their own internal assessment systems without properly considering its negative effects, or without realization of the weak relationships between quantitative indicators and the substantive quality of individual publications, time is pressing.

I. What Problem are Assessment Systems Meant to Solve?

The first question to ask is which problems are meant to be resolved through assessment systems. This article has stated previously that its focus is on the assessment of the quality of scholarly legal publications. However, both peer review procedures and metrics-based assessment schemes are frequently set up to serve multiple objectives. Assessments can provide insights into the quality of scholarly publications and identify opportunities for improvement. In practice however, results from peer review assessments and increasingly, from metrics-based assessments are also used to force researchers to justify their research time in order to obtain funding. Though this may not have solely negative effects, it might lead to conflicts between different objectives. The following example will illustrate this potential for friction.

An objective such as a minimum number of publications per researcher per year (output) requires a totally different incentive than the setting of an upper limit for the number of

⁷² Can and will the editorial board overrule peer reviewers in case of contradicting or poorly motivated reports? What happens with peer reviewers who are very critical, while afterwards the article may be published (elsewhere) with great success? Do journals work with a 'letter to the editor' procedure to allow for correction of peer review failures after publication?

publications over a number of years which can be submitted to reward cutting-edge research and high-impact publications.

The former requires consideration of the fact that many researchers are notoriously self-exploitative and habitually neglect office hours to publish as much as possible,⁷³ out of enjoyment of the end results, because of the importance attributed to the activity, and because it may lead to greater career possibilities, among other reasons. An output-based incentive would therefore saddle a large group with distrust of a small number of stragglers. What is even more important though is that it is unlikely to have a positive influence on the quality of publications because 'scoring enough credits' may become more important than producing innovative results. What also needs to be considered is that groundbreaking research normally is time intensive and has a high failure rate.⁷⁴ All this clearly signals that law school managers need to be patient and have confidence in their researchers, that is at least if they genuinely wish to provide a research environment that encourages high-quality publications.

If the quality of publications does not depend on the publication medium (book, article, essay, etc.), the language, or the type of research (doctrinal, comparative, empirical), but solely on the content of the publications, the community of legal scholars in Europe has an important responsibility to determine which quality indicators are relevant. Further, specific conditions necessary to encourage high-quality scholarly legal publications are also important to set clear expectations for legal scholars. The following sections consider four questions deemed important in this regard.

II. Questions Specific to Different Fields

The first group of questions is linked to the current developments in legal scholarship involving more border-crossing and empirical legal research. Although these types of research in themselves say nothing about the quality of individual publications, a discussion on the constitution of good international or empirical legal research is necessary. After all, there is even less reason to place trust on a tradition of implicit and tacit understanding about quality standards here than in doctrinal legal research.

⁷³ See De Swaan (note 62), 267.

⁷⁴ A famous example is Ludwig Wittgenstein, who was offered a position at Cambridge in 1930 even though he had not published since the twenties and failed to do so until he left again in 1947. It was discovered after his death he wrote 26 pages (on average) on average each day for of those 17 years, but those were for his *Philosophical Investigations*. This was, in turn, published posthumously in 1953 and is regarded as one of the seminal works of Western philosophy. Wittgenstein would probably have been refused a professorship at Cambridge under the purview of the RAE, agrees Donald Gillies in, *Lessons from the History and Philosophy of Science regarding the Research Assessment Exercise*, in *PHILOSOPHY OF SCIENCE* 37 (Anthony O'Hear ed., 2007).

Simply writing in a foreign language, whether or not on an international subject or for an international audience, seems hardly sufficient. After all, many 'international' publications only distinguish themselves from national doctrinal research through their acknowledgements, rather than their approach or reasoning. Cultural philosopher Boomkens has pointed out that an international approach also requires a thorough knowledge of and a lively debate on the national system in order to be able to mirror foreign experiences.⁷⁵ On the other hand, how should we determine if an article about foreign law(s) is not too driven by the political identity and personal beliefs of the researcher?⁷⁶

As far as comparative law research is concerned, an author must justify the choice of countries used as case studies. What should those choices be based on, and how should they be justified in legal publications? Next, what is being compared must be considered—arguments, results, or mechanisms to draw the inspiration of readers? This is followed by the question of whether findings are reliable, and how the results of the comparative study can be applied. Jan Smits sometimes sees normative meaning in its results, whereas Antonin Scalia in the United States categorically rejects any normative meaning of comparative law for answering legal questions that arise in the US system.⁷⁷ These two opposing opinions raise issues regarding imparting advice to starting researchers, and methodologies for subsequent solution-building.

Similar remarks can be made about multidisciplinary research. There the debate is dominated by three questions: (1) where can the information relevant to legal research question in other fields be found? (2) How can the reliability and validity of the gathered information be determined? (3) Assuming the gathered information is reliable and valid, how can its legal meaning be transcribed? For example, what does it mean if empirical research confirms or denies that the death penalty has a deterrent effect (whether temporary or not) on future criminal behavior?⁷⁸ Which *legal* problem does that solve? None of these three questions have a clear answer, partly because answering them requires better and more diverse methodological training, and partly because integration of disciplinary findings is much more complex than most handbooks of law and

⁷⁵ RENE Boomkens, *Topkitch en slow science. Kriek van de academische rede* (Top Kitsch and Slow Science: A Critique on Academic Reason) 87 (2009).

⁷⁶ Arguing that this is a real issue is G. Frankenberg, *Stranger than Paradise: Identity & Politics in Comparative Law*, 259 UTAH LAW REVIEW 259 (1997).

⁷⁷ See SMITS (note 31), 116. For the opinions of Antonin Scalia, compare Melissa A. Waters, *Justice Scalia on the Use of Foreign Law in Constitutional Interpretation: Unidirectional Monologue or Co-constitutive Dialogue*, 12 TULSA JOURNAL COMP & INT'L LAW 149 (2004).

⁷⁸ Compare, Cass R. Sunstein & Adrian Vermeule, *Is Capital Punishment Morally Required? The Relevance of Life-Life Trade-offs*, U CHICAGO LAW & ECON, OLIN Working Paper No. 239 (2005); John Donohue & Justin J. Wolfers, *The Death Penalty: No Evidence for Deterrence*, THE ECONOMISTS' VOICE 3.5 (2006).

corresponding fields suggest. These translation problems somehow need to be reflected in the assessment of scholarly publications, but the way forward remains unclear.

Simultaneously, traditional legal scholars must reflect on doctrinal research, and determine if the information published on a subject does more than just disseminate knowledge. Other scientists often cite that legal scholars 'don't do science, they just argue.' Maybe there is some truth to that, in the sense that the argumentation by doctrinalists is sometimes aimed too much at achieving a 'desirable' result.⁷⁹ Does that not mean that doctrinalists should arm themselves against the reproach that they are not sufficiently open-minded towards normative and empirical reality, and too often indulged in veiled arguments?⁸⁰ If so, what does this mean for the assessment of legal research and the editorial policy of publishers and law journals?

III. Questions Regarding Editorial Policy

The second group of questions that needs answering has been a subject of debate for much longer than the first. Clearly, law journals do not only contain inferior articles and the grass is not always greener on the other side of the fence. The fact that many law journals in the UK, Belgium, the Netherlands and other European countries do not distinguish between scholarly articles and professional or popular publications is an issue. Similarly, the fluctuation of the scientific quality of articles, especially in journals that must cope with a manuscript shortage needs to be considered.

The problem, however, is certainly not that legal practitioners cannot write legal publications with scientific relevance. In that sense the dichotomy between scholarly and non-scholarly publications is false. Neither is it necessary that publications must only be innovative and original while contributing to the body of scholarship. Making legal knowledge more accessible, describing the law on a certain subject, or explaining what the advantages and disadvantages of certain legal decisions are may be very important for legal practice and can have huge societal relevance. However, it does not automatically imply scientific relevance as long as no theory-building is involved, and the publication does not increase the body of knowledge in a certain (academic) field.

Would it therefore not be a good idea to require that law journals themselves separate scholarly and non-scholarly legal publications? Not assigning blame for the present situation, but rather calling on the editors to provide insight into their methods to ensure

⁷⁹ Among others Jan Vranken, *Als een arrest niet bevalt, laten wij het gewoon weg* (We leave out judicial decisions we don't like), in *EX LIBRIS HANS NIEUWENHUIS 77* (Alex Geert Castermans *et al.*, eds., 2009).

⁸⁰ Martijn Hesselink, *A European Legal Method? On European Private Law and Scientific Method*, 15 *EUROPEAN LAW JOURNAL* 20, 22 (2009).

the scientific quality of scholarly publications is a necessity in the success of bibliometrics.⁸¹ It would be quite conceivable for journals that wish to address legal practice *and* encourage scholarly debate to create a separate section for 'scientific publications' and to apply more rigorous and explicit standards to publications appearing in that section. These could include *substantive standards* such as originality, profundity, thoroughness and cross-bordering. Moreover, *procedural safeguards* such as requiring an author to specify the innovativeness of a contribution, setting minimum standards for description versus scholarly analysis, rotating editors to avoid too much 'group think', and forming partnerships with other journals to further exchange of national and international publications and reviewers can also be incorporated into the process.⁸² Experimenting with different sorts of peer review and making the criteria of review more explicit might be useful. However, assuming that this is possible because of the availability of enough capable reviewers could allow alternate solutions involving *methodological justification* to be put into practice..

IV. Questions Regarding the Importance of Methodology

The third group of questions revolves around the importance of methodological justification in journal articles, which have been previously argued by drawing attention to research questions, proper citations, and consistency.⁸³ These questions, and others, surface by themselves where multidisciplinary and empirical research is concerned, but not in doctrinal research. The latter is our primary concern. In the Netherlands a heated debate has ensued over the minimum criteria for methodological justification in scholarly publications, including the need to begin publications with sound research questions.

Although the claims here do not express that an attention to methodology *alone* warrants innovative publications or theoretical profundity, they do imply that this could help editorial boards filter out substandard contributions. Put in another way:

⁸¹ Compare Carel J. J. Stolker, *Legal Journals: In Pursuit of a More Scientific Approach*, 2 EUROPEAN JOURNAL OF LEGAL EDUCATION 87, 89 (2005).

⁸² A scholarship header, such as the one in the *Nederlands Juristenblad* (Dutch Law Journal), would in principle need to be open for good foreign contributions. We would even go one step further by asking if every serious law journal that does not restrict itself to an audience of (national) legal practitioners should open up to contributions from foreign authors? If so, does this not imply that journals with academic ambitions should think about how language barriers can be overcome instead of leaving this entirely to the responsibility of individual authors?

⁸³ Rob van Gestel & Jan Vranken, *Legal Scholarly Papers. Naar criteria voor methodologische verantwoording (Three Methodological Requirements for Legal Research Papers)*, 82 NEDERLANDS JURISTENBLAD 1448 (2007). Available in English on file with the authors.

[T]he techniques, the research community and the methodological rules – together constitute a methodological domain through which all research must pass in order for it to achieve certain standards of integrity and validity. It acts as a mediator between the researcher's subjective beliefs and opinions and the data and evidence that he or she produces through research. If this domain is functioning properly, it acts as something like a filter which prevents bad research from passing through.⁸⁴

However, requiring a proper research question does offer more than merely a filter. It can give the researcher a push towards innovation and original legal research.⁸⁵ A proper research question demands that researchers integrate this into the scholarly debate in a certain field. This may create some distance from practical jurists explaining the theoretical importance of their research by specifying doctrinal positions that are attacked or strengthened, or how certain developments in positive law can be explained or forecast from a multi- or interdisciplinary perspective. This might be done to make explicit research aims adding existing knowledge on the subject, and to explain which evaluative, descriptive or prospective methods among others can be used to answer the research questions, and why the question is deemed suitable for that purpose.

Innovation is based primarily on creativity, intuition, perseverance and the genuine curiosity of the researcher, combined with an open attitude towards the fact that one can always be wrong. This does not imply that legal scholars aiming for originality, profundity, thoroughness alone should not be offered a helping hand. Rather, asking them to make their implicit assumptions, argumentation and assessment framework more explicit could do this.⁸⁶ Provided this does not lead to method fetishism, where for example, the main question of an article must always be labeled as the research question and followed by a number of sub questions, and also provided that the argumentative power is not smothered by formal requirements, scholars and editorial boards aiming for quality can hardly be opposed to this.

What has been said here on the subject of journals is just as valid for other forms of publication such as books, anthologies, essays, case notes etc. Editorial boards or publishers should cooperate with law schools to develop quality standards to filter out scholarship that does not pertain to set standards. The primary motivation for this argument is Pen's research, which shows that performance measurement systems are

⁸⁴ MIKE MCCONVILLE & WING HONG CHUI (eds.), *RESEARCH METHODS FOR LAW* (2007).

⁸⁵ Pushing is not the same as steering. The latter means influencing the substance. Pushing is used in the sense of motivating. On the same subject and on the various functions of a proper research question, see HERVE TIJSSEN, *DE JURIDISCHE DISSERTATIE ONDER DE LOEP. DE VERANTWOORDING VAN METHODOLOGISCHE KEUZES IN JURIDISCHE DISSERTATIE (How to Justify Methodological Choices in Legal Ph.D. Dissertations)* 68, 120, 143 (2009).

⁸⁶ BARBARA E. LOVITTS, *MAKING THE IMPLICIT EXPLICIT: CREATING PERFORMANCE EXPECTATIONS FOR THE DISSERTATION* (2007).

unable to cope with the great variety in research output. Many forms of output are excluded or receive short shift because the system would be too cumbersome otherwise. This appears to be unavoidable, but to quote Pen, is a 'rigid and paltry' system really the price we want to pay?⁸⁷

V. Questions Regarding the Debate on Standards for Legal Scholarship

The fourth and final group of questions asks attention for yet another alternative: Would it not also be conceivable and desirable to shift the attention somewhat from the assessment of publications to an evaluation of legal scholarship? After all, the training and selection of legal scholars probably has a huge impact on the quality of publications.⁸⁸

Sometimes the alarm bell is rung in a certain area of the law, as was the case in the Nuffield report titled 'Law in the Real World'⁸⁹, which denounced the deplorable state of empirical legal research in the UK. The report signaled a lack of research funding and training facilities and a strong focus on professional demands in the law curriculum. In the US the situation is somewhat different.⁹⁰ Empirical legal scholarship (ELS) is flourishing there, at least in the top law schools, but there is also concern about the quality of much of that research.⁹¹

In the Netherlands, there is a debate on the relationship between legal scholarship and legal practice. The aforementioned Stolker committee argued, for example, that legal scholars are sticking too close to legal practice. Scholarly legal debates are mainly result-oriented: is a particular solution better, more just, more efficient, more practical? This is based on the fact that in most legal research, the judiciary is still considered the role model.⁹²

⁸⁷ See Pen (note 2) 151, 153 among others.

⁸⁸ We cannot go further into this here but Osterloh and Frey, for instance, argue that a stricter selection and socialization of academics who want to become legal scholars, for example by using 'tenure track' systems, might seriously downplay some of the negative side effects of quantitative research assessments. Margit Osterloh & Bruno S. Frey, *Research Governance in Academia: are there alternatives to academic rankings?* CESIFO Working Paper No. 2797 26 (2009).

⁸⁹ Dame Hazel Genn, Martin Partington & Sally Wheeler, *Law in the Real World: Improving Our Understanding of How Law Works, Final Report and Recommendations* (2006).

⁹⁰ Tracey E. George, *An Empirical Study of Empirical Legal Scholarship: The Top Law Schools*, 81 INDIANA LAW JOURNAL 141 (2006).

⁹¹ See Epstein and King (note 38).

⁹² JAN VRANKEN, *EXPLORING THE JURIST'S FRAME OF MIND* (2006).

In the US, some argue for a different swing of the pendulum because, as they see it, theoretical work has moved too far away from actual legal practice. Recently, Richard Posner, a well-known adherent of multidisciplinary and empirical legal research claimed that '[T]he law schools *need* legal analysts, not merely as teachers but also as scholars. Doctrinal analysis cannot be left to judges.'⁹³ In a recent book he adds that doctrinal legal research is important for the vitality of the legal system and often of greater social value than much esoteric interdisciplinary legal scholarship.⁹⁴

All in all, there seems to be much disagreement on the quality of legal research and the direction in which the discipline is moving.

G. The Need for a Europe-Wide debate

This disagreement is also reflected in discussions about the assessment of publications. What sort of assessment criteria should prevail? In the UK there seems to be concern among legal scholars about the use of bibliometrics, because some believe that it would push legal research in the direction of the social and natural sciences to a large degree. On the other hand, the Higher Education Funding Council for England proposes just the opposite.⁹⁵ In other countries a similar attention for metrics is starting to appear.

Furthermore, rankings are pressing forward. A popular ranking of legal publications can be found on the Social Science Research Network (SSRN), where one can browse top papers in terms of download numbers. Also available is a list of law schools based on downloads from SSRN's eLibrary, which is updated monthly.⁹⁶ This is not the only private initiative. In Germany for instance, students from Hamburg University developed a ranking of law journals in 2009 on the basis of a survey of 248 professors and staff members of 45 law faculties.⁹⁷ On the other hand, in Belgium and the Netherlands, attempts to establish a ranking of law journals have failed so far as no agreement on the ranking criteria could be reached. The hesitance of editorial boards to clearly distinguish between professional and scholarly publications also played a part in the failure of the system here.

⁹³ Richard A. Posner, *The State of Legal Scholarship Today: A Comment on Schlag*, 97 GEORGETOWN LAW JOURNAL 854 (2009).

⁹⁴ RICHARD A. POSNER, HOW JUDGES THINK 211 (2008).

⁹⁵ Namely to supplement the Research Excellence Framework's (REF) process of peer review. See, *Research Excellence Framework (REF)*, HIGHER EDUCATION FUNDING COUNCIL FOR ENGLAND (HEFCE), available at: <http://www.hefce.ac.uk/research/ref/> (last accessed: 3 January 2011).

⁹⁶ *Legal Scholarship Network*, SOCIAL SCIENCE RESEARCH NETWORK (SSRN), available at: <http://www.ssrn.com/lisn/index.html> (last accessed: 3 January 2011).

⁹⁷ JURAF, available at: <http://www.juraf.de> (last accessed: 8 March 2011).

Considering all this, would it not be useful to have a broader discussion about the assessment of legal research in Europe? The UK is certainly not the only country struggling with the advantages and disadvantages of peer review versus metrics to assess the quality of publications. In the Netherlands the national legal research assessment of 2009 was probably the last truly national and discipline-oriented research assessment exercise. In Belgium, the debate on ranking of journals also seems to have influenced the willingness to begin another national research assessment exercise. The conclusion should thus include a real need for a more uniform system of research assessment. Further, the ranking of law journals demands a European approach in an increasingly international legal research environment. Would it not be wise to use the advantage of the current 'lagging behind' other social sciences to initiate a proper scientific debate on the pros and cons of different systems of quality assessment to ensure uniformity in set-up standards? Perhaps an institution like the European Research Council might lead the way in this matter. Skeptics may be amenable to the fact that an open debate in the current moment could save us a highly politicized battle over harmonization of legal research assessments in the years to come.