A Methodological Improvement in the Evaluation of Research Output: an Adapted use of the Scholarly Capital Model

Completed Research

Michael J. Cuellar
Georgia Southern University
mcuellar@georgiasouthern.edu

Hirotoshi Takeda
Universite Laval
hirotoshi.takeda@fso.ulaval.ca
University of Southern Maine
hirotoshi.takeda@maine.edu

Duane P. Truex
Georgia State University
Dtruex@gsu.edu

Abstract

How the evaluation of research is conducted has significant effects on the field in terms of what work is done, how it is done and who is rewarded. This paper expands on Cuellar, et al (2016) by providing an extended description and critique of the existing method and an overview of its proposed replacement, the Scholarly Capital Model. It shows that the existing method: counting papers in ranked journals uses an under-theorized base, systematically distorted data and has deleterious effects on the field. The paper then overviews the Scholarly Capital Model and shows how it can be used to evaluate research regardless of type of institution.

Keywords

Bibliometrics, Influence, Journal Ranking, Meta-IS Research, Scholarly Capital Model, Social Network Analysis

Introduction

How the research output of scholars is evaluated has significant effects on the field. How work is evaluated determines, in large part, what work is done, i.e., evaluation is itself performative (Mouritsen 2006). Evaluations of research output determine in large part whether scholars receive promotion and tenure (P&T), grants, or awards. Seeking to get positive evaluations of their research, these pragmatic consequences will drive, to a large extent, which research a scholar chooses to commence, how it is done, and the methods by which it is done.

The simple rule-of-thumb ‘count the number of articles published in ranked journals’ approach has evolved into a kind of pragmatic standard for evaluating the quality of academic output. This methodology is attractive as an ‘efficient’ and ‘quick’ method of evaluating scholars within disciplines using an uncomplicated metric. By relying on the knowledge of the journals' review teams who are considered to be experts in the field and are tasked to assess the individual work being submitted, the method avoids the necessity of deans and promotion and tenure committees being expert in all fields and having to read and evaluate every one of a scholar’s papers. While this method seemingly provides a transparent and efficient way of assessing scholars - and the quality of their scholarly output - the current situation is not as clear-cut as it might appear. Criticisms have been made that it promotes the existing forms of research, stifles innovation and incents scholars to work on familiar problems rather than large, complex, and societal ones (Grover and Lyytinen 2015; Winter and Butler 2011).
In this paper, we address these criticisms and argue that the method of evaluating research output by counting publications in ranked journals inaccurately evaluates scholars research contributions because it is atheoretically based, uses bad data, and has deleterious performative effects. In its place, we advocate evaluating scholars based on their scholarly capital, i.e., their demonstrated ability to impact the field. We also reinforce the proposal of a model to operationalize this evaluation approach. Our current paper expands on research by Cuellar et al. (2016a) on the Scholarly Capital Model. Cuellar, et al. provide only a cursory description of the current system and its issues, we provide an extended description and critique of the existing system and then identify how the use of the Scholarly Capital Model, can be used to resolve those issues.

The structure of the balance of this paper is as follows: in the next section, we describe the current system of evaluating research output. We then argue that using the current method does not do what it purports to do and the current system of scholarly evaluation actually distorts the discourse of the field. We then discuss a proposed method of evaluating scholars by means of evaluating their scholarly capital. We then describe how the model can be adapted. We then close with a discussion of implications of this proposal and its limitations.

The Current System of Evaluating Scholarly Output: Journal Ranking

The current system in general use for the evaluation of the quality of scholarly output can be described as “counting articles published in journals.” And it is not sufficient to publish in simply any journal. Each institution has lists of journals which are ranked or stratified into layers of “quality”. The scholar is directed that in order to attain institutional rewards, e.g. promotion, tenure, and/or pay increases that a certain number of publications in journals of a certain quality is expected. This model seems to hold in Australia, France, Canada, the UK and the USA.

A key concept underlying this approach is the journal peer review process provides a kind of warrant to the ‘quality’. The stratification of journals appearing in the ranked journal lists is done by survey or by senior scholars wherein their ranking is assumed to be done based on their informed opinion of the ‘quality’ of the articles published in those journals.

One may argue the current system for evaluating scholarly output has certain advantages and validity. The counting publications in ranked journals approach provides a pragmatic and efficient solution to the problem of how to evaluate scholars from many different fields. By using the journal review process of the field as the arbiter of quality, evaluation teams, who may have been drawn from other academic disciplines, are provided an implicit warrant to the scholar's quality in his/her home discipline, thus alleviating the P&T committees and other reviewers of much of the need to critically examine individual research outputs. This approach saves evaluators time and reduces the necessity to be a subject matter expert in many different fields. The selection of journals by experts allows the most experienced scholars to select and ‘grade’ the venues typically by the creation of journal rankings lists. It is argued that because these 'most experienced' scholars have demonstrated, by their ability to publish in many different venues and make significant contributions to the field, they have the breadth of knowledge to be able to compare journals and come to an informed decision.

A Critique of the Current System

If the present system of research output evaluation has served us so well, then one must ask, “why should the present evaluation system be changed?” Simply stated, the assumptions surfaced above do not stand up to analysis and is based on four spurious/problematic premises. We argue the present method should be changed because: (i) it relies on an undeveloped, implicit theoretical base; (ii) it relies on inaccurate data that results in invalid assessments of the scholar’s work; (iii) and it distorts the discourse of the field. These issues have led to the current valuation process becoming a distorted discourse.

The Concept of Quality is Under-theorized and Implicit

The current method bases its deliberations on the assumption that the field has a valid method for evaluating the quality of publications. This is a dubious assumption at best. The fact that there is no theory of quality or operationalization of quality for academic literature is generally conceded (Dean et al. 2011;
In perusing the literature on evaluation of scholarly quality, we find the construct is constantly referenced, but is used in an implicit and imprecise manner. Quality is not defined and what is meant by the term is assumed to be understood and accepted. This non-definition of quality is defined by Garvin (1984) as the transcendent view. A transcendent view considers quality to be something that cannot be articulated; we know a quality product when we see it but find it hard to pinpoint particular characteristics that make it a quality product. One’s notion of quality is developed by exposure to a succession of objects designated to be of “quality” so we develop a sensitivity for quality. Editors may therefore recognize good research when they see it, as a result of the sensitivity they have developed through exposure to many research publications.

Given there is no generally accepted or even articulated theory of the quality of scholarly output (at best a transcendent concept of quality is applied), and editors’ and reviewers’ notions of quality are developed unsystematically, we can say there is no objective standard to consistently apply to the evaluation of articles. A consistent repeatability of an evaluation across reviewers or across time, also does not exist. Different reviewers may arrive at different decisions, especially when there are multiple philosophical stances that exist, such as methodological positions. Additionally, the same reviewer at different times may have a different perspective. Given this subjectivity the designation of “quality articles” may become, in certain situations, a political decision. Since the concept of quality is largely untheorized and subjective, it leaves the evaluation of the articles subject to the views of quality in the minds of the reviewers thus putting great power in the hands of a few who are relying on idiosyncratic and largely unmeasurable and sometimes unrepeatable decision-making criteria.

The Data Used in the Process is Systematically Distorted

Additionally, we suggest the present method of evaluation is problematic because it relies on bad and/or distorted data about publication quality. In place of explicitly determining the quality of publications offered for P&T, evaluating article quality based on venue of publication assumes articles are of a certain quality because they are published in a particular journal. As will be shown below, this assumption results in bad data for evaluation activities.

First, the data is distorted in that it does not report “quality” correctly. Regardless of what is thought of as quality, the methodology of using publication placement in ranked journals as a proxy for quality results in a distortion of perception. Singh, Haddad, and Chow (2007) reported that in the management field, the top five journals publish only 37% of the most highly cited articles. Of the papers published in the top journals, only 75% were cited more than the average number of citations in the field. They argue that using ranked journals is a very dubious methodology and ought to be discontinued. Similarly, in the IS field (Cuellar et al. 2016b) has shown three different journal lists classify journals into incorrect strata so that the average quality levels are not consistent within the strata; the articles published in a journal or in a strata of journals are not of a consistent quality; articles with a large number of citations can be found in journals classified in the lower levels; and articles with a low number of citations can be found in the higher ranked journals. Thus, indicating that this method results in an erroneous classification of papers. We argue that the current method of using journal rankings saying that high ranked journals publish “quality” articles is flawed. The reasoning here is that “quality” is seen as research with better theory, methodology, and contribution. Here, contribution is the impact of the research on other research and practice. The researchers that evaluate journals and create journal rankings, will be influenced by “quality” research, noting in which journal the “quality” research appears, and create the journal ranking list accordingly. Thus, journal rankings are a proxy for contribution. If this holds true, then citations should be higher for articles appearing in highly ranked journals. But Singh, et. al (2007) and Cuellar, et al. (2016b) did not see this, and found type I and type II errors. Articles appearing in lower ranked journals would routinely garner more citations than those in higher ranked journals, and articles appearing in higher ranked journals would routinely garner fewer citations than those in lower ranked journals. Thus the current method of equating high ranked journals to “quality” articles is flawed.

Second, the data has been distorted by the way publication venues are admitted to the “high quality journals” classification. In some academic settings, an even more divisive discourse has centered on the question of which journals are to be counted. In these settings, advocates of the more selective list choose to ignore the IS field’s own designation of a so called “Basket of journals” as an admissible set of journals that should be considered high ranking when making critical academic evaluations such as P&T decisions.
A Methodological Improvement in the Evaluation of Research Output

(Myers and Liu 2009; Saunders et al. 2009). The overflowing attendance at the panel presentation and subsequent audience discussion during the 2010 ICIS St. Louis conference provided an indication that many in our field are concerned about the notion of “endorsing” a list of journals and about the process by which journals are admitted to the list and thus come to receive a stamp of approval (Information.Technology.Development.Journal 2010). In many cases, the selection of the journals to be included on the list for the institution is the subject to political forces in which journals are added or deleted from the list used in each institution on such bases as who has published there, which editors are on the faculty, other ranking lists, etc.

Deleterious Effects

As if relying on bad data is not bad enough, we note that the current system distorts the discourse in the field by encouraging the wrong sorts of research. As was noted above, Grover and Lyytinen (2015) suggest that the current method of evaluation is very good at encouraging researchers to follow the established model of middle ground research which limits two potentially fruitful avenues for research: those investigations into phenomena without theory and those investigations of theory without data. Their criticism is similar to that of Winter and Butler (2011) who argue that IS researchers prefer to look at small and familiar problems rather than large, complex, and societal ones as was also noted at the Senior Scholars’ panel at ICIS 2012. Similarly, Mingers and Willmott (2013) suggest the evaluation mechanisms in place such as the journal list direct the research conducted, leading toward research designed to be accepted by journals on the list, leading to a lack of creativity, repetition of normal accepted science, and the destruction of creativity and innovation in the field.

The data used to evaluate scholarly output has also been distorted by topical and methodological purity considerations in the review process, which privileges certain types of papers (Smith 2015). As our field has matured, it has struggled to deal with challenges to its own integrity and distinctiveness, particularly with regard to ontological and epistemological openness of research methods and of valid research topics. These challenges have resulted in distortion of the discourse by guiding authors toward specific topics, methodologies, and other practices to ensure publication. For example, prior to 1993, MISQ’s policy was to publish strictly positivist research (Walsham 1995). As a result, those who sought to engage in interpretivist research took a courageous principled stand. Courageous in that by continuing to propagate interpretivist research, they risked their careers by choosing not to be published in MISQ or by being rejected from MISQ. For most however, journal lists established by institutions specified for P&T force scholars into publication regimes favored by those publications.

If one accepts the processes by which we evaluate a scholar might be adversely impacted by bad data resulting from a distorted discourse and adversely impacts the field, then a follow up question must be: how might we generate better data on which to base our decisions? In our view, such a solution proposal should meet four criteria: (i) have a clear theoretical grounding; (ii) have evaluation criteria reducing subjectivity; (iii) have reduced dependence on subjectively derived stratified journal lists; (iv) and be based on measures that are transparent, testable, and reproducible.

Proposed: A Portfolio Approach to Evaluating a Scholar’s Contribution

Instead of the flawed method of counting publications in ranked journals to assess quality, we propose, instead, to substitute the concept of “scholarly capital” which has been defined as “the collection of capabilities and standing that the scholar brings to the organization” (Cuellar et al. 2016a). This proposed basis for evaluation of scholarly output does not attempt to determine the quality of the scholar’s work by reading and somehow evaluating the work, but rather by the uptake of the scholar’s ideas by the field, their connections within the field, and the placement of their publications in the venues of the field. This approach is grounded in a well-developed literature stream in both Lotkaian informetrics (Egghe 2005; Lotka 1926) and social network analysis (SNA) (Freeman 1979) as well as some firm definitions and theoretical reasoning of its own. It follows the ideas proposed by Hassan and Loebbecke (2016) that scientometrics has a role to play in helping to assess and guide the development of the IS field.

We summarize this theory below.
The Scholarly Capital Model (SCM)

Cuellar et al. (2016a) starts by defining the IS research field. “The field” might be defined as the set of scholars who either identify themselves as IS scholars or who hold positions in IS or closely related academic departments (i.e., operations management) but such an approach is likely to be inaccurate, impracticable, and unstable. As an alternative, we propose the IS field is defined by those venues publishing IS research. On the assumption that the content of the articles, or research artifacts, published by those venues are relevant to IS, then any scholar who publishes in those venues may be considered to have contributed to the IS field. The method of defining a field can be directly objectified by using a methodology such as factorial analysis or other statistical classification such as described by Mingers and Leydesdorff (2014).

Having defined the concept of the field, Cuellar et al. (2016a) move on to discuss the concept of scholar capital within the field. The SCM model identifies three forms of capital: ideational influence (who uses your work?), connectedness (who are you working with?) and venue representation (where do you publish?). Together, these three types of influence form the SCM (Figure 1).

Ideational influence is defined as the uptake of a scholar’s ideas by the field (Truex III et al. 2011). To be influential in this manner requires the scholar to have both published work and to have that work referenced.

Figure 1. Scholarly Capital Model (SCM) (Cuellar et al. (2016a,p. 4))
by others. The methods for operationalizing this construct have come from scientometrics. In the IS field, scientometric studies are based on a variety of theoretical groundings and methodological approaches and apply various degrees of scientific rigor (Truex III et al. 2009). The ideational influence construct has been operationalized in the literature by means of citation analysis and the Hirsch family of indices (Cuellar et al. 2008; Truex III et al. 2009; Truex III et al. 2011). The use of the Hirsch index has been shown to be useful in evaluating scholars and superior to other measures such as the impact factor (Mingers 2009; Mingers et al. 2012). In the SCM, ideational influence is operationalized by the use of the h-index (Hirsch 2005), the g-index (Egghe 2006), and the hc-index (Sidiropoulos et al. 2006).

The second form of scholarly capital is connectedness. Connectedness represents the relationships a scholar has in the field which he/she might use to advance the acceptance of her/his ideas. Connectedness derives from the idea that the development of scientific knowledge is well recognized as being a social activity (Bhaskar 1997; Latour 1987; Pinch and Bijker 1984). As researchers work together, they interact with each other to help flesh out theories and test these theories either formally through the publication process, or informally through interactions at conferences and other meetings, or through media such as telephone and email. These interactions formalize into co-authorships. This form of scholarly capital has been operationalized using SNA (Freeman 1979) of co-authorships, reflecting the rationales and consequences of choosing with whom to co-author manuscripts. (Polites 2009; Takeda et al. 2012; Vidgen et al. 2007)

Connectedness is operationalized by use of the betweenness, degree and closeness centrality measures.

Venue representation is the third form of scholarly capital, defined as the kind of resource that arises from the publishing venues in which a scholar’s work appears. An academic field is defined by a set of publication venues - typically journals and conferences - that constitute the methods of knowledge dissemination of that field. The venues as the marketplaces for the discourse of the field serve a function for aggregating what the editors consider interesting, relevant, and important research while screening out less interesting or less-deserving work. Publishing in these venues therefore confers legitimacy to the research findings of academics wishing to contribute to the body of knowledge in their fields and therefore create credibility and ability to influence the field. Venue representation is assessed using the affiliation network, which has a single set of actors (scholars) that are associated with a set of events, in this case the publication venues. In the affiliation network the links are not between the actors but between actors and events (Sasson 2008; Wasserman and Faust 1994), in this case scholars and publication venues. SNA provides tools for analyzing the activity and network position of the scholars in the affiliation network with regard to their publishing activity and their closeness to the journals that comprise the field. Venue representation is operationalized by use of the betweenness, degree and closeness centrality measures.

The three forms of capital are seen as reinforcing (Figure 2). For example, scholars with high levels of connectedness are more likely to have high levels of ideational influence and venue representation. Others might seek out a scholar with high ideational influence for co-authoring opportunities thus leading to higher connectedness. Similarly, he/she might be able to publish in many different publication venues including those most central to the field leading to higher venue representation. By the same token, a scholar with high connectedness might be able to influence having his/her articles read and cited by others. Finally, an author with high venue representation might have his/her papers read more often leading to high ideational influence and receive more invitations to co-author leading to yet higher connectedness. The three forms of capital, then, are codependent and cross influencing supporting positive - and negative - cycles of influence (Cuellar et al. 2016a). When assessing the influence of scholars all three forms of influence should be taken into account: to what extent do others use your work, who do you work with, and how central to the IS field are the venues you publish in?
Utilizing the SCM to Evaluate Research Output

In this section, we describe how the SCM could be used to advise hiring, P&T committees and grant panels. Cuellar et al. (2016a) provides a set of nine measures that constitute a profile of a scholar’s Scholarly Capital. While they showed how to use the SCM in general terms, we are extending this usage by adapting the use to different institutions that value different areas of the SCM. There are three for each of the three components of scholarly capital. An institution preparing to evaluate a scholar’s research output first must define the composition of desired profile for a scholar at that institution. These might be expressed in radar charts as shown in figure 8 from Cuellar et al. (2016a).

The institution will define the value of the levels of each of the three components. A teaching focused institution, for example, will have lower levels than a premier research institution on all of the measures. For example, for promotion to full professor, the institution can develop a target profile based on a scholar at that institution that they wish to replicate or target values based on well-known scholars that they wish to develop. Similar values can be computed for the level of capital desired at the time of tenure. Because the indices used predict themselves well (Hirsch 2007), using the same target scholars used to develop the full professors, their profiles at the time of tenure can be computed and used as targets for the tenure process. In the evaluation process, citation records and co-authorship patterns would be extracted for the scholar under review and then the nine values computed and plotted onto a radar chart as shown. These values can then be compared against the target values. Scholars can then be assessed for meeting the appropriate level of ideational influence, connectedness, or venue representation. Direction could be given for scholars to pursue more leading-edge work, to co-author with different scholars, or to publish in different venues.
A Methodological Improvement in the Evaluation of Research Output

Figure 3: Illustrative Research Profiles (from Cuellar et al. (2016a, p. 20))

Conclusion

We believe that the SCM framework, with its theoretical propositions anchored in established theory and with testable, transparent measures is an evolution and improvement of the method used to evaluate scholarly output because it provides a theoretically based objective set of measures that are replicable, stable, and readily available. By providing multiple measures, it avoids single points of failure and bias. As such provides a credible interim step on route to the development of a more rigorous method of evaluating scholarly output. By using the SCM one can evaluate a researcher by looking at their ideational influence via citation measures (how well is the field taking up your ideas), connectedness by ones co-author network (whom do you work with), and venue representation by their venue affiliation network (where do you publish your work). By doing so the evaluator is looking at a more accurate reading on the researcher’s scholarly capital that they bring to the organization. The SCM in addition alleviates many of the problems of the current system. We believe that it is more difficult to “game”, uses objective standards of measurement and alleviates the “isomorphic” pressures that authors feel to conform to the expectations of reviewers and editors.

To this end it is our hope that this paper will further a critical discourse on how to improve the task of scholarly evaluation and in the end influence not only P&T decisions, as well as the external evaluations that are important to universities, but will also help individual scholars assess the influence of their own work. By examining their profile, they can identify different areas in which they need to focus: increasing
publication frequency, building additional co-authorship relationship or additional venues of publication. Winter and Butler (2011) describe the “grand challenges” in the IS discipline as being: (i) difficult to solve; (ii) demand significant improvements in research; (iii) require great advances of knowledge; (iv) and rely on collaborative efforts from many disciplines and communities. Our hope is that a shift in evaluation can help the IS discipline engage with these “grand challenges”.

REFERENCES


Twenty-fourth Americas Conference on Information Systems, New Orleans, 2018


