Reconsidering Counting Articles in Ranked Venues (CARV) as the Appropriate Evaluation Criteria for the Advancement of Democratic Discourse in the IS Field

Michael J. Cuellar
Duane Truex
Hirotoshi Takeda

Available at: https://works.bepress.com/michael_cuellar/62/
Reconsidering Counting Articles in Ranked Venues (CARV) as the Appropriate Evaluation Criteria for the Advancement of Democratic Discourse in the IS Field

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

Duane Truex  
*Georgia State University*

Hirotoshi Takeda  
*University of Southern Maine*

Follow this and additional works at: [https://aisel.aisnet.org/cais](https://aisel.aisnet.org/cais)

Recommended Citation


This material is brought to you by the AIS Journals at AIS Electronic Library (AISeL). It has been accepted for inclusion in Communications of the Association for Information Systems by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
Reconsidering Counting Articles in Ranked Venues (CARV) as the Appropriate Evaluation Criteria for the Advancement of Democratic Discourse in the IS Field

Michael Cuellar
Department of Information Systems
Georgia Southern University
mcuellar@georgiasouthern.edu

Duane Truex
Department of Computer Information Systems
Georgia State University

Hirotoshi Takeda
Business Analytics and Information Systems
University of Southern Maine
Management Information Systems
Universite Laval

Abstract:

For a research field to advance, scholars must be able to openly exchange ideas. For this open exchange to exist, the contexts and methods that evaluate scholarly output must encourage this interchange. We argue that the current process for evaluating scholarly output, “counting articles in ranked venues” (CARV), creates pressures that result in a distorted discourse inhibiting the growth of the field. We review the current system of evaluating scholarly output and describe its virtues and shortcomings. Then, based on works by Habermas (1984) and Mingers and Walsham (2010), we suggest that the IS field should adopt an improved method of evaluation that enhances the openness of ideas. This method should be objective, reproducible, relatively easy to compute, and standardized. The new method should also produce easy-to-compute profiles of measures that assess a construct for scholarly output founded on a well-defined theory. It should also reduce the dependence on publishing in any particular journal. The scholarly capital model (SCM) represents one such method. We believe that wide adoption of the SCM would open the discourse and help the field to develop more rapidly.

Keywords: Bibliometrics, Critical Social Theory, Democratic Discourse, Evaluation of Research, Habermas, Hirsch Indices, Journal Ranking, Promotion and Tenure, Scholarly Capital Model (SCM), Social Network Analysis.
1 Introduction: Scholarly Advances Require an Open Exchange of Ideas

For an academic field to advance, scholars must be able to freely and openly exchange ideas. Such a free and open exchange requires easily accessible ideas that scholars can openly examine, discuss, and challenge, which, thus, encourages scholarly discussion and further research. An open environment allows scholars to propose, examine, criticize, and revise ideas such that they can identify the most valuable ideas to further develop the field. Conversely, an environment that restricts and filters the flow of ideas such that researchers in the field cannot access them will restrict the field’s development. Blockages or distortions in the discourse limit scholars’ opportunity to see and/or evaluate new ideas that might help advance the field and, thus, delay scholars from recognizing their value. In this debate article, we define the academic publishing process as the activities required to bring research to the field by publication and the subsequent discussions about these publications. We posit that these activities are all part of a discourse that should be open, transparent, participative, and democratic. The notion that a field has a free flow of ideas, which Mingers and Walsham (2010) term a “democratic discourse”, forms part of a larger philosophical domain of discourse ethics and, in particular, Apel’s and Habermas’s notions of “deliberative democracy” (Apel 2001; Habermas 1996; Ketner 2006; Rasmussen 1990; von Schomberg & Baynes, 2002).

Evaluating scholarly output has critical importance in creating a democratic discourse because the ways in which scholars evaluate and reward scholarly output have significant effects on what work they do (i.e., evaluation is itself performative) (Mouritsen, 2006). Evaluation, a form of feedback, provides signals to authors and the field about what is and is not acceptable performance. Evaluation provides a motivation to perform in certain ways such as conforming to normal science and suppressing novel ideas and methods. Thus, the way in which scholars evaluate scholarship impacts the democratic nature of the discourse. Hence, to further develop the field, evaluation mechanisms must seek to minimize distortions in the discourse so that scholars can propose and shift ideas, whether popular or unpopular, received or heretical, established or new. We argue that evaluation criteria must promote four characteristics: openness, participation, truthfulness, and non-privilege.

In many academic settings, the publication represents the “bankable unit” of academic capital. These bankable units are then evaluated via two questions: “how often have you published?” and “where have you published?” (Dean, Lowry, & Humpherys, 2011; Karuga, Lowry, & Richardson, 2007; Lowry, Karuga, & Richardson, 2007; Lowry, Romans, & Curtis, 2004). This evaluative scheme yields a simple count of the number of articles published in journals deemed to be of a certain quality. We term this pragmatic and simply applied standard for evaluating the quality of academic output “counting articles in ranked venues” (CARV). CARV is a proxy for an article’s value such that one imputes quality to articles equivalent to the quality of the venue that publishes them. This rubric represents an efficient shorthand for evaluating scholars in a specific field and for evaluators outside it. By relying on editors and reviewers who others consider experts in the field to evaluate articles, evaluators from outside the field do not need to be experts in many fields or read and evaluate every article that a scholar produces. While CARV provides a seemingly transparent and efficient way to assess the quality of scholarly output—and, thereby, to assess the scholars that produced the scholarly output—the situation is not as clear or simple and unproblematic as it might appear.

Interestingly, scholars have generally recognized CARV as having deleterious and unanticipated effects on any field’s development. For instance, the pressure to publish in journals with “relatively high impact factors”, which leads authors away from new and lesser-regarded journals, has led to types of gamesmanship and outright deceptive practices (Burdick, 2017). For example, many so-called “predatory journals” manufacture fake impact factors (also known as “impact factor rigging”) to boost their reputation. “Citation cartels” work to increase citation counts by citing repetitively in the cartel. The instances of “phony journals”, pseudo-scholarly journals, “false-flag journals”, and “masqueraders” all represent instances of deceptive publishing practices (Anderson, 2015). The attempts to provide false credentials to an author’s article(s) can trick authors into participating, but many critics claim they represent instances of fraud wherein authors complicity participate in the practice (Crotty, 2017). Financial motivations cause some deceptive practices, while inept edit practices in newly created, often “open source”, venues cause others. Regardless, the demand that authors publish in journals with high metrics (even if such metrics sometimes lack veracity) drives all these types of publishing mills.
In this debate article, we examine the criticisms that many IS senior scholars have raised and argue that the CARV method hinders the IS field from developing because it results in constraining democratic discourse. In place of CARV, we posit that scholarly communities need to have a set of measures that will facilitate an open and democratic discourse. Single measure surrogates, such as “top-journal” article counts, will not suffice.

This article proceeds as follows: in Section 2, we describe the nature of a democratic discourse that we seek. In Section 3 we outline how the evaluation method creates performative effects on the publication culture. In Section 4, we outline the current method for evaluating scholars (and acknowledge its strengths). We then illustrate how CARV creates distortions in the discourse and has adverse effects on the field in Section 5. To begin to resolve these distortions, in Section 6, we propose a set of solution criteria that any improved evaluative methodology requires and identify a methodology in Section 7 that meets these criteria. In Section 8, we conclude the article.

2 The Ideal: A Democratic Discourse

In this section, we ask and answer the question: what underpinning principles might an open and democratic discourse adopt? We develop these principles from the features of critical social theory (CST) to explore the implications for assessing research. In particular, we point to Habermas’ theory of communicative action (TCA) (Habermas, 1984), his notion of discourse ethics, and Mingers and Walsham’s (2010) work on developing that idea, which we use as a way to frame a discussion about an open and democratic discourse for the field. We chose Habermas’ TCA because it focuses on the implications of speech and proposes general normative standards for communication. Habermas’ work, and TCA in particular, focuses on “provid[ing] a theoretical framework for the critical analysis of both the structures for and reproduction of discourses of the public sphere” (Cukier, Ngwenyama, Bauer, & Middleton, 2009). Like other scholars who have used Habermas’ TCA to examine the language framing public debate that powerful individuals and institutions use, we recognize critical analysis can inform the reproduction of public and private discourses (Cukier et al., 2009; Mingers & Walsham, 2010; Ngwenyama & Lee, 1997). Indeed, as Culkier et al. (2009, p. 176) states:

[Habermas’] work draws our attention to “the power of these institutions to select, and shape the presentation of messages” and to “strategic uses of political and social power to influence the agendas as well as the triggering and framing of public issues” (Habermas, 1996).

Ideally, we seek a situation in which we can base academic discourse on the notion of deliberative democracy. To describe the nature of a deliberative democracy, we first turn to the basic premises of an ideal speech situation according to TCA. Habermas (1984) identifies two primary types of human action: inherently non-social action versus social interaction. The first orientation has a means-ends rationality and views people as “things” who must comply or be acted on. Alternatively, this orientation may see people as intelligent agents who one can predict or control through language. In Habermas’ terms, the first orientation is referred to as instrumental action and strategic action.

The second orientation concerns interaction with others and has sensemaking and shared agreements as its desired end. Four action types fall into this orientation: 1) normatively regulated action, which depends on shared values and norms with which people will comply; 2) discursive action or argumentation, which deals with the justification of claims and argumentation and actors need when they cannot refer to shared values and knowledge; 3) expressive action, which deals with revealing the nature of an inner-self such as one’s feelings and emotional state; and 4) communicative action, which deals with facilitating understanding and common agreement or consensus. Habermas (1984) considers language use as a

1 Scholars at the AIS Senior Scholars forum at ICIS 2012 came to a general consensus that the current evaluation methodology, whatever its pragmatic virtues, does not represent a good one (Niederman, Kirsch, Lyytinen, Zmud, & Tan, 2012). Winter and Butler (2011) note how IS researchers prefer to look at small and familiar problems rather than large, complex, and societal ones, a preference that the current system of scholarly evaluation reinforces.

2 Jurgen Habermas’ concepts have had considerable recognition and development in the IS literature. We refer the interested reader to works that more fully explicate the backing and genealogy of these ideas. In this article, we use an abbreviated reference to the development of key ideas by Habermas and critical IS scholars to show how these concepts can advance efforts to evaluate scholarly activities. Much research in IS development (Klein, 1988; 1991) and IS research in general (Asif & Klein, 2008; Brooke, Cecez-Kecmanovic, & Klein, 2005; Cecez-Kecmanovic, Klein, & Brooke, 2008; Cukier et al., 2009; Kettner, 2006; Klein & Truex, 1996; Lyytinen & Klein, 1985; Mingers & Walsham, 2010; Ngwenyama, 1991; Ngwenyama & Klein, 1994; Ngwenyama & Lee, 1997; Polites & Watson, 2008; Truex, Cuellar, Takeda, & Vidgen, 2011) uses Habermas’ concepts.
primary form of engaging in human action. Though an inherently social medium, language itself falls into
the two dominant orientations: means-end rationality and shared agreement. The ideal language action
type engenders a rational discourse through which participants arrive at shared understanding and
agreements and then act on those agreements. Forms of discourse oriented to manipulate (strategic
action) or towards a means-ends rationality (instrumental action) create a less effective discourse as they
do not discover truth but argue for a predetermined result. Thus, communicative action is considered the
more ‘ideal’ in advancing understanding and maintaining equality vs. advancing repressive power
differences.

To engage in the ideal open and rational communicative discourse, participants engaged in
communications exhibit a kind of intentional attitude towards fairness and truth. Indeed, Habermas
illustrates this attitude via his three validity claims “truth”, “rightness”, and “sincerity” that, if/when
challenged, a speaker must defend. For Habermas (1984), parties in communication are people who seek
to reach understanding and an informed consensus (even if the people seeking understanding concede
their point of view may not be the final consensus) and they do not seek to get their way via any means
possible regardless of the veracity, strength, or quality of their argument. But Habermas recognized
certain structural and cultural practices might evolve to distort the discourse. For example, applying this
value system to the process of academic evaluation, we take as a given that people engaged in the P&T
deliberative process seek to arrive at accurate and fair decisions—especially since they recognize their
judgments impact the lives and reputations of their colleagues and shape the advancement of
departments, schools, and academic fields.

We recognize, however, that, as Pierre Bourdieu has observed, language is not merely a method of
communication, but also a mechanism of power (Bourdieu, 1984). We also note that powerful interests
may alter current practices and structures to constrict and even close off open discourse. Therefore, ideal
situations remain just that: simply ideal—we are not to look for them in reality but recognize that we strive
toward such situations and that these situations remain the goal of emancipatory action.

Beyond the attitude of goodwill, an ideal speech situation requires four conditions: 1) openness: any party
to a discourse must have an equal opportunity to begin or participate in the discourse and to introduce or
challenge claims; 2) participation: each party must have an equal opportunity to participate in the
discourse; that is, to “interpret, to assert, to recommend, to explain and to justify the validity of any claim
(Klein, 2004, p. 205); 3) truthfulness: the participants must be truthful with each other, seek agreement
and not deception and be equally able to express themselves and police each other in terms of guarding
against self-deceit and insincerity; and; 4) monitored and controlled power differences: differences in
power between the participants will not overtly or covertly restrict participants from openly participating in
the discourse. All participants are free to interact on equal terms with each other, and no discourse is cut
off based on any participant’s whim. When these conditions are met, in Habermassian terms, the
possibility of true argumentation, consensus, and open discourse exists. Consensus represents not only
the presumed outcome of an open discourse but also a goal. One may still have disagreement or
dissensus but have reached it openly and fairly in the realm of the communicative action (Wijnia, 2004).
Whether consensus or dissensus results from the communicative act, space remains for participants to
continue to openly and fairly develop the discourse at a later time.

2.1 Advancing the Deliberative Democracy

Mingers and Walsham (2010) advance the notions of discourse ethics and deliberative democracy for the
IS field by extracting from Habermas’ theory of communicative action what they consider to be
“inescapable presuppositions of rational argument” or core elements and principles required for a rational
and open discourse (p. 840). They conclude: “discourse ethics is intimately related to the TCA because
the domains of the pragmatic, the ethical, and the moral are all embedded within the process of
communicative action” (p. 841). Even though scholars have critiqued Habermas’ notions of ideal
communicative action as being too idealistic and impractical, Mingers and Walsham (2010) build on this
theory saying: “[d]eliberative democracy can be seen to weave together a whole variety of different forms
of discourse and communications involving rational choice and the balancing of interests and ethical
debates” (p. 843). They note that, while complementary, morality and law represent distinct notions. On
the one hand, “morality” is a high-level representation of consensually and discourse-based universally
applicable norms. On the other hand, “laws” enshrine many norms; hence, “the law” must deal with
competing norms and ethical conflicts. Dealing with consensually accepted versus competing-accepted
norms results in three possible changes in the nature of a discourse. First, when dealing with moral
questions, one needs to evaluate issues while considering the interests of all parties. When different communities have similar norms, they can reach consensus. Second, in dealing with ethical questions wherein one faces competing and different norms of different communities, conflict versus agreement then suggests that one cannot generalize to a universal value set. Third, in situations with incommensurate interests between communities and where such communities cannot achieve consensus, one needs bargaining, or other forms of non-rational discourse, such as strategic action. Mingers and Walsham (2010) argue that, even in this third (and least desirable) case, the principles of rationality in discourse can apply to the process if not the outcome:

*The parties involved need to come to a negotiated agreement or accommodation rather than attain a consensus. This is not a rational discourse (in Habermas's terms) since the parties involved will be acting strategically and may well employ power, and because the parties may agree for different reasons, whereas with a moral consensus the parties will agree for the same reasons. Nevertheless, rationality and the discourse principle can be applied to the process of negotiation if not its actual content.* (p. 843)

3 **How Evaluation Affects Field Discourse**

The question may then arise: in what way does the method in which scholars evaluate one another’s published work affect the discourse in the field? We need to address this question because how scholars evaluate work has *performative effects* on the discourse. Evaluation of work is a form of feedback; that is, “actions taken by (an) external agent(s) to provide information regarding some aspect(s) of one’s task performance” (Kluger & DeNisi, 1996, p. 255). Indeed, in several meta-analyses, researchers have identified feedback in classroom environments as one factor that drives improved student performance (Hattie, 1999; Hattie & Timperley, 2007; Kluger & DeNisi, 1996). In the classroom environment, student performance means the student conforms to the instruction paradigm that the teacher provides. Applying this insight to the case of a publication evaluation discourse, the method of evaluation will tend to move the author toward the evaluator’s paradigm. In other words, the reviewing process that the assessment process enforces will motivate authors to conform to what the evaluator prefers. When a P&T process that uses CARV to evaluate the quality of the scholar’s work emphasizes the necessity of publishing in a particular journal, the evaluative process exerts a significant force on authors. This force can result in significant changes to the manuscript and even in destruction of the author’s voice or modification of the submitted article to conform to the wishes of the editors and reviewers (Bedieian, 2004). Therefore, performed the wrong way, evaluation can have destructive effects on the discourse. To maintain a democratic discourse, we want to avoid feedback that carries external compulsion to conform to a particular mode of thought.

4 **The Current System of Evaluating Scholarly Output: CARV**

In Section 4, we describe the current system that academic fields generally use to evaluate the quality of an individuals’ scholarly output as “counting articles in ranked venues” (CARV). For this debate, we assume any university/department P&T body has goals (whether explicit or implicit) for the research that its faculty produces. P&T committees, charged with assessing whether a scholar has achieved an appropriate level, often consult journal lists that rank or stratify journals based on “quality”. If a scholar meets or exceeds the appropriate number of articles in journals of sufficient quality, then such committees deem the scholar to have produced a sufficient quantity and quality of research. The journal stratification itself largely relies on surveys (Hardgrave & Walstrom 1997; Lowry et al., 2004; Mylonopoulos & Theoharakis, 2001; Peffers & Tang, 2003; Rainer & Miller, 2005; Walstrom & Hardgrave, 2001; Walstrom, Hardgrave, & Wilson, 1995) or the opinions of “senior scholars” (AIS, 2015; Cremer, Laing, Galliers, & Kiem, 2015).

Prima facie, one may argue CARV has certain pragmatic advantages and presumed validity. CARV provides a pragmatic and efficient solution to the problem of how to evaluate scholars from many different

---

3 For this debate article, we focus strictly on scholars’ publication outputs when we refer to evaluating scholars. We recognize, however, that, when fully evaluating scholars, evaluators consider other aspects of academic life such as teaching, service, grants, and external letters, etc. We do not hold that evaluators do or should evaluate a scholar solely based on their research publications. Our position has been that evaluators could better assess scholars’ research via a portfolio-of-measures approach (Cuellar et al., 2016a).
fields. By using the journal review process as the arbiter of quality, evaluation teams, who may have come from different academic fields, are provided an implicit warrant as to the quality of the scholar’s work in the scholars’ home field which in turn alleviates the need for the P&T reviewers to critically examine a scholar’s individual research outputs. CARV saves evaluators time and reduces the need that they be subject matter experts in many different fields because actual experts in each field have already selected and ranked journals.

The journal selection and ranking process represents a key concept that underlies CARV. Many evaluators view the ranked journal list as providing a kind of warrant to the “quality” of research because relatively objective experts in a field, a topic, and/or a research method have critically examined it. Further, many assume that the scholars who stratify journals make their decisions based on their informed opinions about the “quality” of the articles that those journals publish. A hierarchy of academic editorial staff who themselves have a reputation as domain experts and who have the responsibility to manage the quality of the review process manage the article-review process. One could argue the evaluation process for whether a journal publishes scholarly work allows for distributed decision making by impartial experts who have the responsibility for ensuring the work, if published, is accurate, well supported, and relevant to the field. One could also argue journals, as aggregators of interesting, relevant, and important research, screen out less interesting or less deserving work and, thus, provide a service to a field and other scholars. The more established and better-regarded publication venues can attract more experienced and more capable persons to serve on their review and editorial teams and, thus, continuously improve or maintain their quality and reputation. The fact that these journals will themselves have higher impact factors and survey rankings should not be a surprise. High rankings, in turn, attract scholars who seek to publish in those journals, which leads to increased competition, higher rejection rates, and better “quality” articles that survive the review process. Therefore, one could make the argument, on average, the best publication venues will publish the best authors and the best work. In this sense, the quality of authors’ work depends on the journals in which they publish their work.

One might also argue CARV has served scholars well for many years. As a field, we have socialized generations of scholars into the system and trained these junior members to exercise communal responsibility by serving as reviewers and serving in editorial positions. The system has represented an evolving consensus among the parties in the evaluation process and directly resulted from a continual discourse in our field about key methodological, theoretical, and topical issues that an academic field requires to develop. However, we ask: at what cost?

5 A Critique of CARV: The Cost

If the present system of scholarly evaluation has served us so well, then one must ask: “Why should we reexamine the present evaluation system? If it ain’t broke why fix it?” We argue that we should reconsider CARV because it 1) at best, only partially measures scholarly output; 2) relies on an undeveloped, implicit theoretical base; 3) relies on distorted data that results in invalid assessments of the field; and 4) has deleterious performative effects on the field. These issues have caused the current valuation process to become a distorted discourse. We discuss each issue in turn.

5.1 CARV is at Best a Partial Measure

CARV represents an attempt to assess an article’s quality to the exclusion of any other characteristics. While some researchers indicate evaluators should consider other qualities, such as impact (Dennis Valacich, Fuller, & Schneider, 2006), when evaluating research, they still focus on determining the quality of the article or developing a methodology to determine quality when discussing such other attributes (e.g., Dean et al., 2011; Dennis et al., 2006; Gillenson & Stafford, 2008; Katerattanakul & Hong, 2003; Lewis, Templeton, & Luo, 2007; Mylonopoulos & Theoharakis, 2001; Peffers & Tang, 2003; Rainer & Miller, 2005; Serenko & Dohan, 2011; Straub & Anderson, 2010; Walstrom & Hardgrave, 2001). Therefore, the review process, which determines whether a journal will publish an article, focuses on evaluating an article based on whether it conforms to the perceived and often implicitly defined standards of “good” research (e.g., adequacy of the literature review; proper adherence to methodological canons; and good, interesting writing techniques). The review process does not have the information necessary to determine what impact the article will make on the field. While one can guess what the field will do with the article, one cannot know for certain at the time of review. The degree to which a field takes up an article’s ideas determines the article’s impact.
CARV is also a “single-measure” approach to evaluation. By their nature, single measures can only evaluate a single part of a construct. To use a single measure flies in the face of current recommendations, Desouza et al. (2007, p. 272) suggest:

_Work to broaden the metrics used in your departments for promotion and tenure. The field is changing, and departments that do not change will be left behind. Action research, ethnographic studies, and even occasionally serving as a consultant or special adviser to an organization, does have merit. Initially, you will be asked to defend your metrics, and this might be difficult, and sometimes even controversial metrics need to measure impact._

Therefore, CARV evaluates only a single characteristic and provides only a single measure of that item; thus, it constitutes a partial measure. Accordingly, one cannot use CARV to obtain the complete perspective of a scholar's contribution. While CARV violates several conditions of the ideal speech situation, we argue that, due to this incomplete perspective of a scholars’ contribution, it violates the truthfulness condition in particular.

### 5.2 We Lack Adequate or Explicit Theory for “Quality” in CARV

CARV bases its deliberations on the idea that one has defined and can measure the quality of publications—a dubious assumption at best. Researchers generally concede the fact that no theory of quality or operationalization of quality for academic literature exists (Dean et al., 2011; Locke & Lowe, 2002; Straub & Anderson, 2010). In perusing the literature on evaluating scholarly quality, we found that, while researchers constantly reference the construct, they use it in an implicit manner without definition. In other words, researchers assume that others know and accept what “quality” means. Garvin (1984) defines this view as the transcendent view: that is, a view that considers quality as something that we cannot articulate; that we know a quality product when we see it but find difficulty in pinpointing particular characteristics that make it so. In Pirsig’s (1974, p. 106) words, this view considers quality as “neither a part of mind, nor…a part of matter. It is a third entity which is independent of the two” and (p. 91) “Even though Quality cannot be defined, you know what Quality is.”. In other words, the field is relying on the construct as being “a usefully ambiguous term”.

The transcendent view of quality deals with quality aesthetics, and one can liken it to primitive Platonic concepts such as “beauty” and “truth”; by being exposed to a succession of quality objects, we develop sensitivity for quality. Editors may indeed recognize good research when they see it due to the sensitivity they have developed through exposure to many research publications. Indeed, many editors take this approach. For example, Straub and Anderson (2010, p. x) indicate:

_[L]et us suggest that a concept like journal quality lies almost completely in the minds of scholars because quality itself is highly abstract…. Without clearly mapped physical markers, we can come up with a set of metrics that will approximate this construct, but never tap into it without a large dose of humility…. It is not even remotely similar to the construct of something physical like ball bearing quality, where we can measure with small degrees of precision the variances of machine tools in creating the balls, their housings, and the processes that assemble these._

Given that we lack a generally accepted or even articulated theory of the quality of scholarly output (at best reviewers apply a transcendent concept of quality), which puts more power in the hands of the reviewers who rely on idiosyncratic and largely unmeasurable decision-making criteria. In this sense, the current method fails the standard of truthfulness. If we do not have a standard on which to judge, how can we maintain truthfulness?

### 5.3 CARV Relies on Systematically Distorted Data

Additionally, we suggest CARV is problematic because it relies on data that reviewers derive from a systematically distorted discourse. In place of explicitly determining the quality of publications, the present method assumes articles have a certain quality because a particular journal publishes them. As we show in this section, this assumption results in distorted data.

First, the assumption results in distorted data because it does not report “quality” correctly. Regardless of what one regards as quality, the methodology of using journals as a proxy for quality results in a distorted perception. Singh, Haddad, and Chow (2007) report that, in the management field, the top five journals publish only 37 percent of the most highly cited articles. Of the articles that the top journals published, only...
75 percent received more citations than the average amount of citations in the field. Accordingly, they argue that using ranked journals for evaluation is a dubious methodology and ought to be discontinued. In the IS field, researchers have found journal-ranking lists to lack a consistent level of quality. That is, they have found that journals found in a particular strata in ranking lists to vary widely in quality according to the number of citations they received. Additionally, the lists often misclassify articles. In a previous study (Cuellar, Truex, & Takeda, 2016b), we examined the capability of journal lists to properly classify articles. In that study, we examined journals based on citation counts to determine if the articles published in the journals should be in the strata that they appeared in. We analyzed articles from journals in each stratum to see if they fit in each stratum or not. We found that type 1 errors (articles that should not have been in each strata of a ranking list) occurred over 43 percent of the time, while type 2 errors (articles not classified in the correct strata) occurred over 35 percent of the time (Cuellar, Truex, & Takeda, 2016b). Thus, we can conclude that journal lists do not have much value as identifiers of article quality.

Several possible reasons explain this finding. The process is neither fully open nor fully participative because all parties cannot equally participate in it. Since journals have space limitations, journals of a particular level lack the space to publish all the articles at their level (e.g., articles that match a particular quality level). Researchers have documented the scarcity of “valued venues” well (Adams & Johnson, 2008; Dennis et al., 2006; Katerattanakul, Han, & Rea, 2006; Mylonopoulos & Theoharakis, 2001). Most major journals have page count restrictions on their publications, which limits the number of articles they can publish. These restrictions limit the publishing opportunities in those journals. Journals may delay or even reject articles that they could and should publish due to space considerations. While top-tier journals have increased the number of articles published in recent years, the number of scholars seeking publication has also increased at a greater rate, which has resulted in more difficulty in achieving publication in top-tier journals (Dennis et al., 2006). Anecdotal evidence indicates that top-tier journals must contend with an excessive number of submissions, while lower-level journals often scramble to find suitable articles to publish. This scarcity situation violates the openness and participation condition of an ideal speech situation.

Second, the assumption results in distorted data due to the way publication venues become designated as “high-quality journals”. In some academic settings, an even more divisive discourse has centered on the question of which journals one should count. In these settings, advocates of the more selective list choose to ignore the IS field’s own designation of the so-called “Senior Scholars’ basket of journals” as an admissible set of journals that one should consider to have a high ranking when making critical academic evaluations such as P&T decisions (Myers & Liu, 2009; Saunders, Brown, Sipior, Zigurs, & Loebebecke, 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 have concerns about the notion of “endorsing” a list of journals and about the process by which journals join the list and, thus, come to receive a stamp of approval. In many cases, political forces influence what journals institutions select for their journal lists such that they add or delete journals from the lists based on who has published there, which faculty members serve as editors, other ranking lists, and so on.

Third, topical and methodological purity considerations in the review process distort the data, which privileges certain types of articles. 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 distorting the discourse by guiding authors toward specific topics, methodologies, and other practices to ensure publication. For example, prior to 1993, MISQ had a policy to publish strictly positivist research (Walsham, 1995). As a result, those who sought to engage in interpretivist research took a courageous principled stand—courageous because, in continuing to propagate interpretivist research, they risked their careers since MISQ would reject and/or not publish their work. For most, however, journal lists that institutions establish for P&T decisions force scholars into publication regimes that those publications on the list favor.

Since CARV does not measure what it purports to measure (i.e., quality), it violates the condition of truthfulness, and, since it often does not allow different methodologies to enter the discourse at times, we argue that CARV violates the openness and participation conditions.

5.4 CARV has Performative Effects on the Field

The use of CARV will inevitably have effects on the field as it is the de facto method of measurement for academia. Researchers have generally recognized CARV as having deleterious effects on the IS field’s development. Bedeian (2004) argues the review process constitutes a negotiation that socially constructs
the text based on the interaction between the authors and the reviewers such that the authorial voice is often lost. The reward system creates pressure that pushes authors to bow uncritically to the reviewers’ demands in order to publish their work in the journals that they need to publish in to achieve a promotion or tenure. This pressure violates the power difference condition that an ideal speech situation requires. MacDonald and Kam (2007) suggest the evaluation system creates gamesmanship in both authors and journal editors in that both try to maximize their placements. Mingers and Willmott (2013) argue that using journal lists forces authors to conform to editors’ and reviewers’ desires, which results in a conformance to normal science and, thus, a “taylorization” of research topics and methodologies. Finally, Grover and Lyttinen (2015) argue the review process results in a forced conformance to a “script” and that, while good at getting articles published, it forces research into a middle ground model that privileges theory application at the expense of articles that explore data without theory and develop theory without data. These issues result in closing the discourse, which hinders the field from developing. None of these critiques suggest that the negotiation between authors and reviewers that constitutes the review process enhances an article’s quality.

Additionally, one can argue the present reward system, which recognizes only articles published in “top” journals, stifles the development of the field by not rewarding valuable activities such as reviewing and editing. Most evaluation systems consider these activities as a “service”; that is, as a minor consideration compared to teaching and research. As such, scholars receive little credit for reviewing and editing and, thus, have an incentive to minimize their involvement in this activity, which constricts the amount of journal and conference reviewing available and, in turn, scholars’ opportunities to publish.

5.5 CARV is the Antithesis of an Open and Democratic Discourse

Given our analysis, we posit that the four issues we mention in Sections 5.1 to 5.4—partial measurement, implicit standards, distorted data, and performative effects—violate the necessary conditions for an ideal speech act and, thus, violate the standards for an open and democratic discourse. The first three conditions (partial measurement, implicit standards, and distorted data) violate the truthfulness standard in that we deal with data that does not reflect the actual situation. The final condition (performative effects) results from power differences in the process that distort the communications that scholars provide to the field. Those scholars who cannot publish in ranked venues established by their institutions cannot access the discourse and the rewards of the academic system, which violates the openness and participation conditions. These issues result in a closed and non-democratic discourse among the field’s members.

If one accepts that a distorted discourse might adversely impact the processes by which we evaluate scholars, then we need to ask: how might we create conditions approximating a more ideal discourse? In our view, such a solution proposal should also conform to open discourse conditions and should have a clear theoretical grounding; evaluation criteria to reduce subjectivity; a reduced dependence on subjectively derived stratified journal lists; and transparent, testable, and reproducible measures.

6 Solution Criteria

Thus far, we have argued that we need an open and democratic discourse to advance the field and, thus, a research evaluation mechanism that promotes such discourse. In examining CARV above, we can see that it does not promote a democratic discourse. Thus, we need to ask what better evaluation methodology we could use? One could promote many such solutions. In order to evaluate them, we need a set of evaluation criteria. Based on our criticism of CARV, we offer the following set of solution criteria.

6.1 Assess More Important Attributes

As we determine above, scholars typically use quality in an undefined and subjective manner. We suggest that quality is an inherently subjective concept difficult to define and operationalize. In its place, the construct measured must be one that acknowledges a scholar’s ability to do research and add knowledge to the field. One such construct might be influence or impact. Influence or impact measures the ability to move the field’s thinking by means of the contributed ideas (Truex et al., 2011).

6.2 Any Replacement Methodology Must be based in Sound Scientific Theory

Many scholars have criticized CARV as an expedient measure that lacks any theoretical foundation that explains why it is a good measure. Any assessment methodology is an information-processing artifact. As
such, the artifact should be based on some sound scientific theory that provides evidence that it will perform the function it should perform (Simon, 1996). In Simon’s thought, artifacts are created to perform a function in a particular environment. He argues that, for an artifact to work, the artifact’s functional mechanism must be based on sound scientific theory and that the design of the artifact must consider the nature of the environment in which it will operate. In this case, any new evaluation methodology should establish a set of metrics that will accurately evaluate a chosen criteria that assesses research. Such methodology should be based on established scientific theory that predicts that we will accurately evaluate research on our chosen criteria.

6.3 Any Replacement Methodology Must Provide a Profile of Metrics Rather than a Single Measure

Evaluation should be based on a balanced and more open and egalitarian set of measures. Researchers recognize that single measures do not give a broad view into a phenomenon. Therefore, we need multiple measures to give multiple views. Thus, when evaluating a scholar’s research output, we need to have multiple measures that correspond to multiple views into their performance.

6.4 Any Replacement Methodology Must be Relatively Easy to Compute, Objective, Reproducible, and Relatively Easy to Standardize (CORS)

CARV possesses one outstanding attribute: its ease of computation. To find a methodology simpler than counting articles published in ranked journals would be difficult. As such, any replacement process should strive to be relatively easy to compute. An automated measurement process would fulfill this need. One could simplify even a complicated process that involves large amounts of data (e.g., authors’ bibliometric data and the field’s bibliometric data if one processes analyzed network effects) with an automated method of computing the metrics. In addition to ease of computation, it should be more objective. That is, it should not involve any subjective analysis. As we point out above, CARV involves much subjectivity. While counting publications is objective, ranking journals and reviewing criteria are essentially subjective processes. We recognize that any method that humans develop is inherently subjective, but using a process published in an algorithm at least gives transparency and reproducibility. Such a method has more objectivity in that scholars clearly know it and can validate any assertion based on it by rerunning the algorithm. A replacement process should also be reproducible. Reproducibility allows verification against error and fraud. Scholars could resolve any disputes about the measure by recomputing the measures. The ability to automate implies the methodology would be relatively easy to standardize. A standard automated routine that collects and processes the data would provide ease of computation, objectivity, reproducibility, and standardization across organizations. The CORS standard will provide a way to compare all proposed methodologies.

6.5 Any Replacement Methodology Should De-center the Journal and Journal Lists

As we saw above, the CARV methodology emphasizes journals as the method to evaluate articles, and thus, scholars. We describe how this emphasis has raised various extrinsic motivations that have resulted in deleterious effects that have distorted the discourse in the field. To resolve this issue, we argue processes that evaluate scholars’ research capabilities should not be principally predicated on their publication in an exclusive and limited set of journal venues. Rather, such processes should use criteria that avoid authors having to change the novel ideas in their articles to conform to editors’ and reviewers’ wishes or to conform to established methodologies and practices.

7 A Possible Methodology to Replace CARV

One such proposal has been the scholarly capital model (SCM) (Cuellar, Takeda, Vidgen, & Truex, 2016a). Rather than attempt to assess quality, this model proposes that one assess the concept of scholarly capital, “the collection of capabilities and reputational assets that a scholar brings to an organization” (academic department, research group, or other organization based around research) (Cuellar et al., 2016a, p. 4). The SCM argues that scholarly capital comprises three subconstructs: ideational influence (the extent to which the field takes up a scholar’s ideas), connectedness (a scholar’s ability to draw on the scholar’s relationships with other scholars to advance the scholar’s ideas), and venue representation (the ability to publish in venues more central to the field). Cuellar et al. (2016a) operationalize SCM with a portfolio of nine measures. Three measures (h-index variants) assess
ideational influence. Another three measures (SNA centrality measures) come from a social network analysis of co-authorship data and measure connectedness. The final three measures (centrality measures from the analysis of an affiliation network of scholars and venues and measure venue representation. As a comparative measure, SCM takes on evaluative meaning only when one compares scholars. To use it in practice, an evaluator needs to establish standards for the different measures of the portfolio. One can then evaluate scholars against those standards. The SCM is also customizable. Each institution can establish the levels of ideational influence, connectedness, and venue representation that they wish to see in their scholars by focusing in on the three measures of each construct. Academic units could then proactively use the approach to plan and shape hiring and promotion decisions based on the portfolio-of-skills approach similar to the analytical and sabermetric approach as in the book (and subsequent movie) Moneyball (Lewis, 2003), which leads baseball teams to assemble winning teams by choosing players who have specific focused strengths versus vying to fill each position with the best player they can afford.

The SCM meets the requirements for an improved method of analyzing a scholar’s research by assessing scholarly capital rather than some poorly defined and imprecise notion of “quality”. Indeed, the SCM proposal has clearly defined and operationalized scholarly capital so that it is not subject to varying interpretations. The SCM builds on sound scientific theory. The measures used in the portfolio are the products of a long stream of research. One can find the h-index variants in the bibliometric research stream that originated with Lotka (1926) and progressed through Egghe (2005) to the foundational articles by Hirsch (2005), Egghe (2006), and Sidiropoulos, Katsaros, and Manolopoulos (2007). The social network and affiliation networks build on the research that stems from Freeman (1979), Wasserman and Faust (1994), Sasson (2008), Brieger (1974), Borgatti and Everett (1997), and Borgatti, Everett, and Freeman (2002). To fulfill the requirement of multiple metrics versus a single metric, the SCM provides a profile of nine metrics that measure three different constructs rather than a single number. The SCM also fulfills the CORS standard. It is easy to automate and easy to compute. One could construct an algorithm to extract the citation and co-authorship data and produce the nine statistics for the SCM. Since SCM relies on data rather than subjective opinion surveys, it has more objectivity. Such an algorithm, if open to all and, thus, transparent can be openly assessed and known and, therefore, is more objective than lists of ranked journals based on subjective standards in evaluators’ heads. Since one can automate it, one can reproduce the statistics as required. Since it is easy to compute and automate, one could relatively easily standardize the calculation methodology. Finally, since it uses h-index statistics, social network statistics, and affiliation network statistics, the SCM does not focus on journals and, thus, decreases the deleterious effects from the review process.

8 Conclusion

Do we wish to advance the IS field more rapidly by encouraging a democratic discourse that enables scholars to fully consider all ideas? In a real sense, we diminish all researchers if we allow a distorted discourse. Based on our deliberations, we believe that we need to improve the current system of assessing a scholar’s influence, although well established and familiar, due to its insufficiencies. Rather than a one-dimensional assessment of “quality” with journal rankings, we provide solution criteria that will open the discourse. A proper solution methodology should be an easy-to-compute, objective, reproducible, and standard profile of measures that assesses a construct other than quality. This methodology should be founded on a well-defined theory that justifies why the mechanism employed in the methodology properly assesses a scholar. This methodology will also reduce the dependence on the journal. We believe that, by meeting this set of criteria, a replacement methodology will open the field to more ideas and provide a more democratic discourse. The scholarly capital model constitutes one possible replacement methodology that we believe meets these criteria. If widely adopted, we believe that the discourse would become more open and democratic and the field would develop more rapidly. We look forward to a vigorous debate on how to use research evaluation to best create a democratic discourse.
References


Peffers, K., & Tang, Y. (2003). Identifying and evaluating the universe of outlets for information systems research: Ranking the journals. *Journal of Information Technology Theory and Application, 5*(1), 63-84.


About the Authors

Michael J. Cuellar is an Associate Professor of Information Systems at Georgia Southern University. His research focuses on the areas of project management and organizational change, critical realism as applied to Information Systems and the process of research evaluation. He has substantial experience in industry having held management positions for EDS and American Software managing software product development and infrastructure services, as well as outsourcing sales. He has published papers at conferences and journals such as the *Journal of the Association for Information Systems, European Journal of Information Systems, the European Journal of Operational Research*, as well as *ICIS, AMCIS, and the International Research Workshop on Project Management*. He is a member of the Association for Information Systems and the Project Management Institute. He is Editor-in-Chief for the *Journal of the Southern AIS* and a Senior Editor for *JISE*.

Duane Truex holds joint appointments in the Computer Information Systems Department and in the Institute of International Business of the J. Mack Robinson College of Business at the Georgia State University where he also serves as the Program Director for the GSU’s University of Nantes (France) Academic Exchange program. He has held additional academic appointments in Sweden (as Professor of Industrial Economics at the Mid Sweden University, in France (as a Research Professor at Université de Nantes) and in England as a former Leverhulme Fellow (at Salford University). In addition to his inquiries into the nature of scholarly influence, his research explores the emergent and performative properties of language as instantiated in information systems development (ISD) and in the design of enterprise systems, the effect of organizational emergence on systems architectures and post-implementation governance of enterprise-wide systems (ES), and the social impacts of information systems (IS) on society.

Hirotoshi Takeda is an Assistant Professor of Business Analytics and Information Systems at the University of Southern Maine in Portland, Maine, and Associate Professor of Management Information Systems at Université Laval in Quebec City, Canada. He has seven years of industry experience in telecommunications, semiconductor manufacturing, and IT consulting. After his career in industry, he graduated with a PhD in Computer Information Systems from Georgia State University and a PhD in Management from the University of Paris Dauphine. He has degrees in electrical engineering and computer science from UC Irvine, a Masters of Electrical Engineering from the Georgia Institute of Technology and his MBA from Southern Methodist University. He was formerly a post-doc researcher at the University of Nantes, France, and was a adjunct professor at North Carolina Central University. He has also been a visiting professor at Aarhus University in Denmark. His research interests include discourse analysis, mobile computing, bibliometrics, virtual communities, knowledge management, supply chain management, and green IS. His research has appeared in the *Journal of the Association for Information Systems, European Journal of Information Systems, Information Systems Educators Journal*, and the proceedings of the ICIS, AMCIS, SAIS, UKAIS, ISECON, and IFIP WG 8.2.