Do Peer Reviews Predict Impact? Evidence from the American Sociological Review, 1978 to 1982

Misha Teplitskiy¹ and Von Bakanic²

Abstract
The authors investigate how well peer reviews of articles published in the journal American Sociological Review between 1978 and 1982 predict the articles’ citation impact in the following 32 years. The authors find no evidence of a relationship between review outcomes and citation impact at any time after publication, even when citations are normalized by subfield. Qualitative analysis of the review texts rules out the interpretation that reviewers focused on potential impact but failed to predict it. Instead, reviewers focused on the soundness of the manuscripts’ arguments. The authors discuss how organizational characteristics of review can decouple reviewers’ judgments from impact.

Keywords
bibliometrics, citations, evaluation, impact, peer review

Peer review is a crucial mechanism of resource allocation in science. Scientists and institutions depend on the judgments of peer reviewers to distribute billions of dollars and the limited pages of academic journals. It is thus important to understand how reviewers reach their decisions and whether the decisions are reliable and valid. Although the typical confidentiality of peer review has greatly constrained research, the literature has nonetheless grown to include hundreds of studies (Bornmann 2011 reviewed recent advances). The great majority of these studies focus on reliability: to what extent reviewers agree with one another (Bornmann, Mutz, and Daniel 2010; Cicchetti 1991; Cole, Cole, and Simon 1981). Validity of the decisions has proven a much more difficult target. Here we focus on the validity of manuscript review decisions in sociology.

Peer review is commonly conceptualized as the statistical problem of inferring the true value of a scientific claim from a noisy signal (Lee et al. 2013; Mervis 2015). To assess the validity of decisions regarding a claim, one must then identify a claim’s true value, or at least to define the characteristics of the evaluation process that produces unbiased value judgments (Lee et al. 2013). Yet in the social sciences and humanities, “value” may take on a variety of meanings. For instance, good work may be that which is competently conducted, or uses data creatively, or asks a novel question (Guetzkow, Lamont, and Mallard 2004; Lamont 2009). In interpretive disciplines scholars may define good work as that which influences a conversation (Lamont 2009:61, 72). True value may thus be a multidimensional concept, and evaluating whether peer reviewers “get it right” necessitates specifying a dimension of interest, or at least how the various dimensions are to be made commensurable (Lee 2015).

One dimension of value, a claim’s potential for impact, is of special interest. First, it is often explicitly desired by funders and publishers of science. For example, Michael Lauer, a director with the National Heart, Lung, and Blood Institute, “explicitly tell[s] scientists [that potential for impact] is one of the main criteria for review” (quoted in Mervis 2014:596). Second, impact is a unique dimension of value in that it is also a reward and can thus promote a virtuous cycle. In settings in which rewards are reputational, such as academia (Bourdieu 1975), an excellent reputation requires visibility, which is shaped largely by academic journals (Clemens et al. 1995). Journals bring work to the attention of peers, and if peers cite the work, they make it yet more visible, incentivizing more such work in the future. It is thus desirable that the several dimensions of value correlate (are

¹University of Chicago, Chicago, IL, USA
²College of Charleston, Charleston, SC, USA

Corresponding Author:
Misha Teplitskiy, University of Chicago, Department of Sociology & KnowledgeLab, 1126 East 59th Street, Chicago, IL 60637, USA
Email: misha@uchicago.edu
rewarded) with impact. Third, impact is plausibly and easily measured by the classic, if beleaguered, metric of citations. Although scientists’ motivations for citing the existing literature vary (Bornmann and Daniel 2008; Nicolaisen 2007), citations are widely considered a measure of academic impact (Mingers and Leydesdorff 2015; van Raan 1996).

A number of studies in the life and natural sciences have used citation impact to evaluate review decisions. The logic is attractive: for many tasks, decisions of a crowd are more accurate than decisions of a few individuals, so perhaps citing decisions of a crowd of scientists represent a suitable gold standard against which to evaluate (the few) reviewers’ decisions (Bornmann and Marx 2014; Lee et al. 2013). The present study extends this line of research for the first time to sociology, in which evaluation is among the least understood steps of the research process (Leahey 2008).

To answer this question we use the review files of American Sociological Review (ASR), the flagship journal of American sociology, and focus on 167 research articles published between 1978 and 1982 (Bakanic, McPhail, and Simon 1987). The age of the data allows us to evaluate how well peer reviews predicted citations over virtually the articles’ entire lifetimes: 32 years following publication.

Although impact is an appealing gold standard of quality, in settings in which value is multidimensional, it is important to understand the criteria that motivate citers’ and reviewers’ decisions. In practice, analysts do not measure these criteria and assume them instead. For instance, it is often assumed that the audience of potential citers and reviewers decides whether to cite or to review positively, respectively, using identical criteria; in contrast, we focus on the outcome of citers’ behavior, regardless of motivations, and measure reviewers’ criteria. We ask the following questions:

**Research question 1:** Do reviews of published articles predict citations?

To answer this question we use the review files of American Sociological Review (ASR), the flagship journal of American sociology, and focus on 167 research articles published between 1978 and 1982 (Bakanic, McPhail, and Simon 1987). The age of the data allows us to evaluate how well peer reviews predicted citations over virtually the articles’ entire lifetimes: 32 years following publication.

Although impact is an appealing gold standard of quality, in settings in which value is multidimensional, it is important to understand the criteria that motivate citers’ and reviewers’ decisions. In practice, analysts do not measure these criteria and assume them instead. For instance, it is often assumed that the audience of potential citers and reviewers decides whether to cite or to review positively, respectively, using identical criteria; in contrast, we focus on the outcome of citers’ behavior, regardless of motivations, and measure reviewers’ criteria. We ask the following questions:

**Research question 2:** What dimension of quality do reviewers focus on? Do they prioritize potential impact?

To answer these questions we turn to the texts of the reviews, analyzed previously by Bakanic, McPhail, and Simon (1989). By assessing whether reviewers focus on impact or other dimensions of value, we are better able to interpret the relationship between reviews and citations. In particular, if reviewers fail to predict impact, we can adjudicate whether they prioritize impact but fail to predict it, or if they focus on dimensions of value that are uncorrelated with impact.

The rest of this article is structured as follows. First, we review existing studies that use citations to evaluate peer review validity. Second, we describe our data and method. Third, we present our results and discuss the possible interpretations. We finish with a discussion of policy implications.

**Peer Review, Validity, and Citations**

**Reliability and Validity**

Despite the common metaphor of peer review as a “black box,” hundreds of studies have followed Robert Merton and colleagues’ pioneering effort to investigate peer review processes (Merton and Zuckerman 1971). The majority of this literature focuses on reliability: how much reviewers of a particular manuscript or grant application agree with one another (Bornmann et al. 2010; Cicchetti 1991). Studies of the validity of peer review are rarer, nearly absent in social science. For instance, of the 18 studies of grant review identified by van den Besselaar and Sandström (2015), nearly all focused on the life and natural sciences. The rarity is unsurprising: evaluations of validity require not only the reviewers’ decisions but also a metric against which these decisions can be compared.

**Methodological Trade-offs**

**Accepted versus Rejected.** Using citations as the gold standard of article quality requires trade-offs. In the case of academic journals, citations are usually observed only for the manuscripts judged publication worthy. Negatively judged manuscripts either disappear from purview altogether or reappear as articles published in other (usually lower impact) journals (Bornmann 2011:219). If initially rejected articles accrue fewer citations after being published elsewhere, is the cause their lower quality or the publishing journal’s lower status? Or, if citations increase after publication elsewhere, perhaps the articles were revised in a way that increased citability?

In his review of the scientific peer review literature, Lutz Bornmann (2011:218–25) found only five studies that compared citations received by manuscripts published and rejected by particular journals and six studies comparing citations to papers from funded and rejected grant proposals. These five journal-based studies all found that rejected manuscripts receive fewer citations than accepted ones (p. 222). However, none of these studies took into account that rejected manuscripts, if ultimately published, are generally published in lower impact journals. The findings from the grant and fellowship competitions were contradictory, which Bornmann speculated is due to the more prospective nature of these competitions (p. 230). Calcagno et al. (2012) surveyed a population of authors regarding submission patterns and obtained data on 80,748 articles published in 923 biomedical journals. They found that rejected articles ultimately published elsewhere received significantly more citations than articles accepted by the first journal. It is unclear if these results were biased by a low response rate (37 percent). In another study, Siler et al. (2015) found that articles desk-rejected at top biomedical journals and published elsewhere received fewer citations than articles rejected after review and published elsewhere.
Accepted versus Accepted. The approach taken here focuses only on the accepted manuscripts. This approach avoids the problem of variable journal prominence but suffers from limited variation in review scores: only high-scoring items are published or awarded. Figure 1 illustrates the consequences of limited variation.

In a setting ideal for evaluating journal review decisions, articles judged above and below a publishability threshold would both be published. Articles below the threshold, the light red dots in Figure 1, would then generate citation counts just like articles above the review score threshold, the dark red dots. Of course, in practice, the analyst observes only the dark red dots. A regression line among these (the black line) may or may not indicate a relationship. However, if all articles were observable, that is if the observed and hypothetical counts were considered together, a relationship between their scores and citations could be quite pronounced (the dark gray line). This could be the case if expert reviewers can make only gross distinctions in quality but fail to make finely grained ones, as is commonly believed (Mervis 2014). Furthermore, the true relationship could be nonlinear: below a (high) publishability threshold, review scores are associated with citations, but above the threshold, gradations in reviews are of no import.

Considering only published articles can nevertheless yield a valuable and valid interpretation for articles exceeding a minimum threshold of judged quality. The generalizability of conclusions to the review process as a whole naturally depends on the variation in review scores observed: the higher the variation in quality among accepted manuscripts, the better the results should generalize to all reviewed manuscripts.

The evidence from “accepted versus accepted” studies has been mixed. Siler et al. (2015) used data from elite medical journals and found no statistically significant association between review scores and citations of published articles. Danthi et al. (2014) examined the relationship between percentile rankings of 1491 National Heart, Lung, and Blood Institute grants and 16,793 publications resulting from these grants. The authors found no relationship between grant percentile and citation metrics; however, a machine-learning algorithm was able to identify citation “hits” from review scores.

In perhaps the largest study of its kind, Li and Agha (2015) found that among publications generated from the 137,215 National Institutes of Health (NIH) grants, an increase of 1 standard deviation in percentile ranking was associated with 17 percent more citations. In another study of the NIH, Park, Lee, and Kim (2015) used as a natural experiment the unexpected federal stimulus funds in 2009 that enabled the NIH to fund some proposals that had been initially rejected. They found that initially rejected proposals produced 6.8 percent fewer citations per month than initially funded (higher rated) proposals. In sum, results from studies of peer review validity are mixed. In most cases, initially rejected articles receive fewer citations than accepted ones, but the impact of the publishing journal on citations is generally not taken into account. In some cases, there is not a relationship between citations and the review scores of accepted articles or grants, but in two important studies, there were statistically significant, albeit small, relationships.

By and large, the aforementioned studies have assumed away the motivations of reviewers and citers and interpreted the association between review scores and citations as an unproblematic referendum on the effectiveness of peer review as a whole. In contrast, we take into account the criteria used by the reviewers and develop additional interpretations. We argue that the extent to which reviewers are representative of the audience and what the reviewers take as their mandate can both mediate the relationship between reviews and citations.

Data and Methods

Between 1971 and 1981, 2,337 manuscripts were submitted to *ASR*. A sample of these manuscripts and their review data were originally used to evaluate several aspects of *ASR*’s review process, including fairness and reliability (Bakanic et al. 1987, 1989; Simon, Bakanic, and McPhail 1986). Details of the data collection and coding process may be found in Bakanic et al. (1987:634–35). Here we focus on the 167 published full-length articles. Additionally, Bakanic et al. (1989) analyzed the content of the reviews of 323 published and rejected manuscripts. We return to these content analyses here to assess to what extent reviewers focus on the potential for impact.
Reviewer Selection and Instructions

To select reviewers, the editorial staff considered three primary criteria. The staff first looked for reviewers whose work was in the same substantive area and, upon identifying suitable individuals, took into account the quality of their previous reviews (if any) and turnaround time. They excluded people who were known coauthors with or students of the author(s). They avoided (if possible) sending a review to a colleague in the same department and/or institution (C. McPhail, personal communication, January 7, 2016).

Reviewers were mailed a cover letter from the editor’s office, a copy of the manuscript, and a review form with instructions and a list of possible recommendation (reject, revise and resubmit, accept conditional, and accept). The instructions requested reviewers to comment on the appropriateness of the manuscript for ASR, as well as the quality of the literature review, theory, methods, analysis, conclusions, organization, writing, and any other aspect of the paper they thought appropriate. The editorial staff met weekly to consider the manuscripts with completed reviews. They also discussed manuscripts that for which they had trouble finding reviewers (C. McPhail, personal communication, January 7, 2016).

Review Outcomes: Consensus and Average Review Score

Reviewers’ recommendations in a particular round were dichotomized as the consensus variable, taking on the value “consensus accept” when all reviewers chose “accept” or “conditional accept” on their referee forms and “no consensus” when at least one reviewer chose “revise and resubmit” or “reject.” In one case, the editor overrode reviewers’ unanimous decision in the last round of “revise and resubmit.” We also converted reviewers’ recommendations into numerical scores using the scale 1 = reject or refer, 2 = revise and resubmit, 3 = conditional accept, and 4 = accept. Each article’s average review score is the mean (numerical) recommendations made by reviewers in the last round of review.

First versus Last Round. Most manuscripts underwent two or more rounds of review, with review scores assigned in each round. Consequently, the analyst faces a choice between using scores assigned in the initial round of review or those given in a later round. Authors revise manuscripts between rounds, and it is not unusual for the changes to be substantial (Teplitskiy 2015). Reviewers’ opinions of the final manuscript may have little to do with their opinions of its initial form, and it is the final version that accrues citations. Consequently we chose to focus on reviewers’ scores from the last round (an analysis of citations vs. first-round review scores may be found in Appendix Figure A1). On the other hand, scores given by a particular reviewer across rounds are generally constrained to increase. A reviewer who realized that his or her initial score was mistakenly favorable would find it very awkward indeed to explain to the editor and authors why, after the authors undertook revisions, a score worsened (and why the reviewer was not thorough enough the first time). In sum, last-round scores have the merit of a tighter connection with the final article and the demerit of smaller variation (biased upward) in scores.

Figure 2 summarizes the distributions of review outcomes.

In the left panel of Figure 2 is the number of (published) articles in each round and consensus type. In the right panel of the figure is the distribution of (published articles’) average review scores.

Total Citations

Citations received by each manuscript in each year after publication were obtained from the Thompson Reuters Web of Science database between May 23, 2014, and February 17, 2015. Total citation counts include citations received by each article within the first 32 years after publication. Figure 3 displays a histogram of total citations.
The histograms of total citation counts are greatly left skewed; the obvious outlier, with 1,204 citations, is Cohen and Felson’s (1979) article “Social Change and Crime Rate Trends: A Routine Activity Approach.” Tables A1 and A2 in the Appendix list the 10 most and least cited articles in the data sample. We address the skew of citations in the following analyses by log-transforming counts or using nonparametric measures (i.e., ranks) and statistical tests (i.e., Kolmogorov-Smirnov [KS]).

Subfield-normalized Citations

Research fields vary in how frequently typical items are cited. It is thus one of the “key principles of citation analysis that citation counts from different fields should not be directly compared with each other” (Waltman 2015:21). Sociology is a notoriously heterogeneous discipline (Smelser 2015), and ASR welcomes and publishes material from all of its subfields. To compare citations across these various subfields, we normalize citations by subfield in two ways. First, following common the bibliometric practice (Waltman 2015), we define for each article a reference set that consists of other articles on the same substantive topic. The topics consisted of the 54 sections of the American Sociological Association (ASA) and 9 additional topics. Coders classified each manuscript as belonging to up to 3 of these topics (e.g., “medical” or “theory”). All manuscripts sharing at least one of these subfields were included in the reference set.

Normalized citations carry information not present in the raw citations: the two quantities correlate weakly (0.13), as is apparent in Figure 4, which displays their scatterplot.

Results

In this section we examine the relationship between reviews and (un-normalized) citations first and then use normalized citations. Next, two analyses probe whether especially high citations are associated with the highest scores. Finally, review texts are used to argue that reviewers focus on the soundness of manuscripts’ arguments rather than their potential for impact.

Review Scores and Citations

Figure 5 presents a scatterplot of citations versus average review score (in the last round) with a line of best fit. (A similar scatterplot with review scores from the first round of review may be found in Appendix Figure A1). No relationship is apparent visually or statistically between the reviews and citations: the slope of the regression line is $-0.098$ ($SE = 0.187$, $p = .601$).

Review Scores and Normalized Citations

Does the apparent lack of a relationship between reviews and citations reveal a null relationship, or is a small but significant true relationship swamped by other determinants of citations, particularly target audience size? After all, an outstanding article (high review scores) targeted at a relatively small community may receive fewer citations than a mediocre article (low review scores) targeted at a relatively large community. Unmeasured variation in audience sizes may introduce a large amount of noise in citation counts and make all but the strongest relationships appear null. To take into account audience size, citation counts were subfield normalized by dividing citations by the mean citations received by other articles on the same substantive topic. Figure 6 displays a scatterplot and line of best fit for these (total) normalized citations versus review scores.
The relationship in Figure 6 echoes that of the previous figure: no relationship is apparent between review scores and normalized citation counts.

**Multivariate Analysis**

Bivariate analyses displayed in Figures 5 and 6 show no evidence of a relationship between review scores and citations. However, the bivariate relationships did not control for important predictors of citations identified or posited by the literature. These predictors include the number of coauthors (Wuchty, Jones, and Uzzi 2007), institutional prestige (Leimu and Koricheva 2005; Long 1978), and subfield spanning (Leahey and Moody 2014). To control for these covariates, we estimated the following regression:

\[
\text{NORM.CITES} = \beta_0 + \beta_1 \text{REV.SCORE} + \beta_2 \text{NUMB.AUTH} + \beta_3 \text{AUTH.PRESTIGE} + \beta_4 \text{AUTH.RANK} + \beta_5 \text{SUBFIELDS} + \epsilon
\]

NORM.CITES (normalized citations) and REV.SCORE (average review score) were defined earlier. AUTH.PRESTIGE (first author prestige) was measured as the prestige of the first author’s current institution (1 = high school, 5 = MA-granting institution, 10 = PhD-granting institution with an American Council on Education ranking of 60 or higher. AUTH.RANK (first author rank) was measured as the first author’s professional rank (1 = undergraduate, 5 = research associate, 9 = professor emeritus). SUBFIELDS was measured as the number of substantive research areas the manuscript addressed (see note 11).

Table 1 displays estimates of this regression.

The central quantity of interest—the regression coefficient of the average review score—is statistically insignificant. Meanwhile, the number of coauthors is the strongest and most reliable predictor of citations. This multivariate analysis provides further support to the earlier bivariate analyses: even when controlling for important covariates, the data do not provide evidence of a relationship between review scores and citations.

**Do Peer Reviewers Identify Citation “Hits?”**

If reviewers do not predict the impact of accepted articles overall, perhaps they do successfully identify citation “hits”: the articles that accrue an uncommonly large number of citations (Danthi et al. 2014). If reviewers and editors recognize that an article with great potential is before them, they may wish to make its path to publication as quick as possible and accept it after the first round of review. In our data, there are 18 such “first-round consensus-accept” articles. Comparing citations received by this group—the “cream of the crop”—with citations received by the 149 articles taking other paths to

---

**Figure 5.** A scatterplot of citations versus average review score for each manuscript (red dot) and a line of best fit (black). The correlation is −0.051. The slope is −0.098 and is statistically insignificant at the .05 level. The dots are jittered to improve visibility.

**Figure 6.** A scatterplot of normalized citations versus average review score for each manuscript (red dot) and a line of best fit (black). The correlation is 0.040. Citations were normalized by dividing each article’s citations by the mean citations received by all other articles on the same substantive topic. The slope is 0.14 and is statistically insignificant at the .05 level. The dots are jittered to improve visibility.

**Table 1.** Coefficient Estimates from the Regression of Normalized Citations on Manuscript Covariates.

| Variable                  | Coefficient | 95% Confidence Interval |
|---------------------------|-------------|-------------------------|
| Average review score      | −0.23       | −2.56 to 3.07           |
| Number of authors         | 0.74***     | 0.20 to 1.28            |
| First author prestige     | −0.030*     | −0.065 to 0.004         |
| First author rank         | 0.038*      | 0.002 to 0.073          |
| Number of subfields       | −0.49       | −1.18 to 0.19           |

Note: Number of complete cases = 106, \( R^2 = 0.16 \).

***p < .05, **p < .01.
publication would then indicate whether unusually favorable review outcomes correspond to unusually high citations.

The distribution of citations is skewed, as noted previously. To ensure that results are not unduly influenced by one or two outliers, we convert articles’ raw citation counts into ranks: the article receiving the least citations is ranked 1 and the article receiving the most is ranked 167. It is then possible to observe whether one of the two article groups—first-round consensus-accept articles versus all other articles—is overrepresented in the low or high article ranks. Figure 7 displays these densities using kernel density estimates plots.13

Figure 7 demonstrates that that first-round consensus-accept articles are not overwhelmingly represented among the high article ranks. In fact, the small overrepresentation among the most highly cited articles is balanced by a small overrepresentation among the least cited articles.

A statistical test confirms the visual evidence. The KS test is a nonparametric (unaffected by skew) test of whether two probability distributions differ in any part of the distribution. A KS test of whether the distribution of (1) first-round consensus-accept articles’ citations differs anywhere along the distribution from (2) all other articles’ citations does not reject the null hypothesis of no difference (KS statistic = 0.17, \( p = .73 \)).14 We thus find no evidence that review “hits” are associated with citation “hits.”

**Impact at Each Year after Publication**

The age of the data invites an analysis of whether review scores predict not only accepted manuscripts’ total citations but also citations at each year after publication. The manuscript were divided into two groups: consensus-accept manuscripts are those that all reviewers recommended for acceptance, while no-consensus manuscripts are those at least one reviewer recommended be rejected or revised and resubmitted. Citations were obtained for each year after publication. Medians were used instead of means to limit the influence of one outlier (using means does not qualitatively affect the conclusions). Figure 8 displays the median citation trajectories of these consensus-accept and no-consensus groups.

The visual evidence is clear: the two groups of articles, one with superior reviews and one with inferior reviews, have nearly indistinguishable citation trajectories over the entire 32-year period.15 The variance of both groups dominates the minor (and unstable) differences in their medians. Thus, the apparent independence of reviews and citations 32 years after publication established earlier holds throughout the observation period.

**Do Reviewers Value Citation Impact but Fail to Predict It?**

All of the results above point unanimously to the conclusion that review outcomes of accepted articles do not predict their citation impact. The lack of a relationship raises a question: do reviewers value citation impact but fail to predict it, or do they privilege other manuscript characteristics, such as soundness of argument? The review instructions offer little guidance; as discussed in the “Data and Methods” section, reviewers were asked to comment on the “appropriateness” and “quality” of the manuscripts. These rather open-ended reviewer instructions are not unusual for sociology. Furthermore, sociologists’ formal training rarely includes review instruction, making reviews practices difficult to deduce from the instructions and potentially nonuniform. As one sociologist put it, “Manuscript reviewing must be one of

![Figure 7. Articles are split into two groups on the basis of review outcome—first-round consensus-accept articles versus all others—and the presence of each group across the range of ranked citations is plotted. The densities shown are kernel density estimates.](image1)

![Figure 8. Solid lines indicate cumulative median citation trajectories of consensus-accept and no-consensus articles. Review decisions are from the last round of review. Shaded areas represent 25th- to 75th-percentile trajectories of each group.](image2)
the most important, least formally trained professional functions that we serve” (Brunsma, Prasad, and Zuckerman 2013).

The focus of reviewers’ attention has usually been elicited with surveys and interviews (e.g., Bornmann, Nast, and Daniel 2008; Brunsma et al. 2013; Chase 1970; Lamont 2009). A chief problem with eliciting by asking, in addition to desirability effects, is that the reviewers are not forced to compromise between criteria in the way they must when actually reviewing. There is no realistic constraint forcing respondents to trade off among criteria: they may simply respond that they value them all. These considerations motivate our choice to deduce reviewers’ foci of attention from the actual review texts. But before we turn to these texts, consider what reviews would focus on if citations truly were of utmost importance.

Expert reviewers are likely to have in mind mechanisms that generate citations in their field. Indeed, impact in the field is to some extent a prerequisite of being the type of reviewer ASR would call upon. Some reviewers may even be familiar with the voluminous scholarly literature on the predictors of impact. If mandated to predict citations, reviewers would likely use all this knowledge in evaluating a manuscript. They would hesitate to accept a well-executed piece of scholarship if it were unlikely to generate citations, perhaps because an earlier article solidified the first-mover citation advantage (Newman 2009) or because it did not address several audiences (Leahy and Moody 2014). A manuscript from poorly positioned authors or research areas could also drag citations down (Bornmann et al. 2012) and would thus be a legitimate cause for rejection. Articles presenting useful methods, which are overrepresented among the most cited articles, would win praise easily (Van Noorden, Maher, and Nuzzo 2014). And of course reviewers could write directly, “I am afraid this manuscript will not be cited.”

The contents of the ASR review reports did not mention citations explicitly, and even implicit mentions were exceedingly rare. The vast majority of review text in these data, whether comments to the author or editor, whether praise or criticism, is devoted to general impressions and to concerns about the soundness of a manuscript’s argument (Bakanic et al. 1989).

Positive comments were relatively rare and, when made, were most often generic (36.8 percent of all comments) (Bakanic et al. 1989, Figure 2). For example, reviewers praised manuscripts by calling them “interesting” (Bakanic et al. 1989:644) or writing “On the whole this is a nice article” (p. 645). Reviewers did often (19.9 percent) praise a manuscript’s topic. For example, one reviewer wrote, “This is an important topic for studies of subjective class identification” (p. 645). Such comments on topic may be indirect endorsements of the manuscript’s citability. In sum, even if style (9.7 percent) and topic (19.9 percent) are taken to be endorsements of citations, nearly 70 percent of reviewers’ positive comments focused on either very general impressions or the argument’s soundness.

Reviewers’ negative comments to the editor were most often general (31 percent of all criticisms). These included statements such as “This is not sociology,” “I wasn’t convinced by this research,” even “Yuk!” (Bakanic et al. 1989:646). Reviewers provided to the authors more concrete criticisms, focusing on problems of theory (13.2 percent), analysis (12.3 percent), and results (11.7 percent) (p. 647). Criticisms of the manuscript’s importance to the field, which may be construed as related to its citability, are relatively rare: 6.4 percent of comments to the author and 6 percent of comments to the editor criticized the manuscript’s topic. Similarly, criticisms of style, perhaps also related to citations, constitute 10.8 percent and 11.9 percent of comments to the author and editor, respectively.

The texts of reviewers’ reports thus indicate that reviewers do not focus on impact. They do not mention citations or other measures of impact explicitly at all; and if impact were understood to be part of their mandate, how could they not? Furthermore, comments that may implicitly relate to citations (i.e., regarding a manuscript’s topic) are much rarer than comments regarding the soundness of argument. Instead, reviewers take as their chief mandate to evaluate the soundness of a manuscript’s argument, addressing most comments to matters of theory, measurement, and analysis.

In summary, ASR reviewers did not make invalid predictions of impact. Instead, reviewers assessed the soundness of manuscript, and for accepted manuscripts, these judgments are unrelated to citations.

Discussion

In this study we explored the relationship between accepted manuscripts’ peer review scores and citation impact. In addition to a possible linear relationship between scores and raw citations, we examined citations normalized by subfield, citation “hits,” and citations at each year in the 32-year span after publication. In none of these specifications did the data manifest a relationship. Did reviewers value potential impact but fail to predict it? Nearly all previous studies did not measure reviewers’ motivations directly and interpreted reviewers’ ability or inability to predict citations as a referendum on the validity of peer review (Bornmann 2011; Mervis 2015). In contrast, in this study we used the texts of the reviewers’ reports to deduce the foci of their attention. The texts show that reviewers did not prioritize potential impact directly or indirectly; instead they focused on the soundness of arguments. Reviewers’ judgments of soundness, not potential impact, are unrelated to citations.

Although this study helps rule out one interpretation of the absence of a relationship between scores and citations, several interpretations remain. First and foremost, the true relationship between review scores and citations may be nonlinear: there may be a threshold of quality below which a strong relationship exists but above which it disappears. This may be the case where the chief distinction (threshold) in
quality lays between rejected and accepted (whether immediately or after revisions) manuscripts, while gradations in quality among the accepted articles are inconsequential (Mervis 2015). Two additional interpretations have been largely overlooked by the literature and are discussed below.

**Nonrandom Reviewer Selection**

Reviewers are not selected randomly from a population of potential citers, so their judgments may fail to generalize to this population. Both reviewer solicitation and agreement to review predispose the final reviewer panel to consist of individuals who find the manuscript’s research area valuable. First, editors at *ASR* and likely in other settings solicit reviews from experts in the field from which the manuscript comes. These experts would not have made the enormous investments required to master their specialties if they did not find the specialties valuable. Second, research on motivations of volunteers in crowd-sourcing settings emphasizes that volunteers donate time in order to learn and be creative (Lakhani and von Hippel 2003; Lakhani and Wolf 2003). We suspect that in scientific peer review, these motivations are equally crucial: experts agree to perform the time-consuming and uncompensated task based, in part, on interest in the research area. In sum, reviews come from individuals uncommonly interested in the manuscript’s research area. Even superlative assessments from reviewers may thus fail to translate into citations from the broader population if it has no interest in the manuscript’s topic.

**Implicit Division of Labor between Reviewers and Editors**

Conspicuously absent in our discussion and others in the literature are the editors. Editors may play a crucial role in mediating the relationship between reviews and citations not only because they select reviewers but because the reviewers may rely on an implicit division of labor between themselves and the editors. In particular, some reviewers may take it as their mandate to evaluate manuscripts for soundness and leave judgments of whether it is interesting, judgments that may be strongly related to citations, to the editors. This implicit division of labor is apparent when one experienced sociological reviewer, in a survey of ASA reviewers, wrote, “a fourth test [of the manuscript under review] is whether the point is interesting or innovative, but I feel that this can be left for the editor to decide so I don’t make a big deal of it” (Brunsma et al. 2013:2). However, in the absence of formal reviewer training or explicit instructions, it is not surprising that reviewers vary in what they take to be their task. For example, in the same ASA survey, another reviewer wrote, “I simply read [the manuscript] from the perspective of a potential reader. The standard and maybe only thing I look for is this: Is this paper interesting? That is my prime criterion” (p. 9). In sum, reviewers differ in their evaluative criteria and in how they divide the evaluation task between themselves and the editors. To the extent that reviewers expect editors to judge a manuscript’s interestingness, the absence of a relationship between accepted manuscripts’ reviewer scores and citations is unsurprising: the crucial relationship to evaluate is between citations and editors’ decisions.

**Conclusion**

In this study we focused on the evaluation of sociological manuscripts, a crucial and understudied research practice (Leahy 2008). We returned to the decisions made by *ASR* reviewers and editors more than three decades ago to assess whether the decisions predicted the published manuscripts’ citation impact over what is essentially their entire lifetimes. The chief empirical finding is a null one: reviews do not predict citations in a number of plausible specifications. In addition to extending the empirical base of the literature to sociology, the study developed a number of interpretations of the relationship especially relevant to social science, where the value of manuscript excellence can be multidimensional. By combining bibliometric and qualitative data, the number of these interpretations of sociological peer review was reduced. In particular, ruled out is the interpretation that reviewers prioritize impact but fail to predict it. Yet several interpretations remain. The relationship between review scores and impact may be nonlinear, with an association at lower levels of quality and no association above a publishability threshold. Additionally, reviews may fail to predict citation because of nonrandom reviewer selection or because an implicit division of labor assigns the difficult task of predicting citations to the editors.

The age of these data enabled us to trace the fate of manuscripts through an unprecedented length of time: 32 years, virtually the articles’ entire lifetimes. Yet the age may limit applicability to contemporary academic publishing. It is possible that the bibliometric conception of impact prevalent today would have been distasteful or foreign to reviewers 35 years ago. The standardization and quantification literature suggests that rating and ranking systems do not simply quantify existing reality; they also induce evaluated and evaluating persons to behave in ways that fall in line with the system of valuation undergirding those systems (Espeland and Sauder 2007; Espeland and Stevens 2008). The increased prevalence and ease of using bibliometrics to measure impact may have changed scholarly norms for assessing the value of work. It is thus plausible that reviewers today, more conditioned to think and evaluate in terms of bibliometrics, may be more successful forecasters.

Despite the limitations above, this study contributes to our understanding of peer review and policies that may improve it. One direction for improvement concerns reviewer instructions. If a journal prioritizes citation impact, reviewers should be explicitly instructed to evaluate on this...
dimension. Another direction concerns reviewer selection. In the case of *ASR*, the two types of a manuscript’s “evaluation,” that by expert reviewers and that by a more general citing audience, appear decoupled. It is plausible that more representative review panels will result in higher associations between review outcomes and citations. To the extent that citation impact is valued, editors may need to trade domain expertise for representativeness.

**Appendix**

**Table A1.** The 10 Most Cited Articles in the Data Sample.

| Author(s) | Year | Title | Citations |
|-----------|------|-------|-----------|
| Lawrence E. Cohen and Marcus Felson | 1979 | “Social Change and Crime Rate Trends . . .” | 1,202 |
| Ronald L. Akers et al. | 1979 | “Social Learning and Deviant Behavior: A Specific Test . . .” | 476 |
| Delbert S Elliott | 1980 | “Reconciling Race and Class-differences in Self-reported and Official . . .” | 410 |
| Nan Lin, Walter M. Ensel, and John C. Vaughn | 1981 | “Social Resources and Strength of Ties: Structure . . .” | 402 |
| Karen Cook | 1978 | “Power, Equity and Commitment in Exchange Networks” | 394 |
| David A. Snow, Louis A. Zurcher, Jr., and Sheldon Ekland-Olson | 1980 | “Social Networks and Social Movements: A Microstructural Approach . . .” | 394 |
| Thomas A. Heberlein and Robert Baumgartner | 1978 | “Factors Affecting Response Rates to Mailed Questionnaires . . .” | 390 |
| E. M. Beck, Patrick M. Horan, and Charles M. Tolbert II | 1978 | “Stratification in a Dual Economy: A Sectoral Model . . .” | 385 |
| Paul D. Allison | 1978 | “Measures of Inequality” | 367 |
| Kenneth A. Bollen | 1980 | “Issues in the Comparative Measurement of Political Democracy” | 281 |

**Table A2.** The 10 Least Cited Articles in the Data Sample.

| Author(s) | Year | Title | Citations |
|-----------|------|-------|-----------|
| Richard P. Applebaum | 1978 | “Marx’s Theory of Falling Rate of Profit . . .” | 0 |
| Brian L. Pitcher | 1978 | “Diffusion of Collective Violence” | 0 |
| Ronnie Steinberg Ratner and Paul Burstein | 1980 | “Ideology, Specificity, and the Coding of Legal Documents . . .” | 1 |
| David E. Payne | 1978 | “Cross-national Diffusion: The Effects of Canadian TV . . .” | 2 |
| John W. Gartrell | 1981 | “Inequality within Rural Communities of India” | 3 |
| Israel Adler | 1978 | “Cross Pressures During Socialization for Medicine” | 3 |
| Colin Campbell | 1982 | “A Dubious Distinction? An Inquiry into the Value and Use of Merton’s . . .” | 4 |
| G. Edward Stephan and Douglas R. McMullin | 1981 | “The Historical Distribution of County Seats in the United States . . .” | 6 |
| Michael C. Burrrage and David Corry | 1981 | “At Sixes and Sevens: Occupational Status in the City of London from . . .” | 6 |
| William C. Rau | 1980 | “The Tacit Conventions of the Modernity School . . .” | 6 |

**Table A3.** Pairwise Correlations between the Variables Used in the Multivariate Regression.

|                      | Total Citations | Normalized Total Citations | Average Review Score | Number of Authors | Number of Subfields | First Author Prestige | First Author Rank |
|----------------------|-----------------|-----------------------------|----------------------|-------------------|---------------------|-----------------------|-------------------|
| Total citations      | 1.000           | -                           | -                    | -                 | -                   | -                     | -                 |
| Normalized total citations | 0.128       | 1.000                       | -0.013               | 0.013             | 1.000               | -0.013                | 0.489             |
| Average review score | 0.034           | 0.040                       | 1.000                | -0.013            | -0.202              | 0.195                 | 0.067             |
| Number of authors    | 0.211           | 0.013                       | 0.013                | 1.000             | -0.077              | 0.007                 | -0.090            |
| Number of subfields  | -0.077          | -0.202                      | 0.052                | -0.195            | 1.000               | 0.007                 | 0.489             |
| First author prestige| -0.016          | 0.002                       | 0.100                | 0.067             | -0.013              | 1.000                 | -0.090            |
| First author rank    | 0.003           | 0.007                       | -0.011               | 0.073             | -0.090              | 0.489                 | 0.000             |
Slightly jittered to improve visibility.

SE statistically significant (p = 0.14, r = .13). The dots have been of review. The slope of the regression line equals 0.22 and is not

Research on revisions is limited, and the impact of revisions

1. Prominent recent examples include Li and Agha (2015) and Siler, Lee, and Bero (2015).
2. Reviewers may also fail to accurately assess these other dimensions. Our data do not allow us to address this possibility.
3. For recent reviews, see Bornmann (2011), Lamont (2009), and Lee et al. (2013).
4. Research on revisions is limited, and the impact of revisions on citation is unknown. Goodman et al. (1994) found that revisions reduce “spin.” Decrease in spin may be expected to decrease citations. Two studies of sociology journals found that revisions alter the framing of results (Strang and Siler 2015; Teplitskiy 2015), but the implications of change of frame on citations are unclear.
5. Bornmann and Daniel (2008) came closest to properly taking publishing journals’ Impact Factors into account by comparing published and rejected-and-published articles’ citations with a baseline of all journals within the discipline of chemistry or one of its subfields. Unfortunately, this approach does not preclude the possibility that journals publishing initially rejected manuscripts are still of higher Impact Factors than the field’s average.
6. It does not appear that the authors accounted for the possibility that desk-rejected articles were resubmitted to lower impact journals than articles rejected after review.
7. Five percent of incoming manuscripts were desk-rejected or withdrawn by the author(s). Of the remaining manuscripts, 63 percent were rejected after review, 25 percent were requested to be revised and resubmitted (and 19 percent eventually published), and 9 percent were unconditionally accepted (Bakanic et al. 1987:634).
8. This set includes nearly all of the articles published between 1978 and 1981 and some published in 1982. We exclude from analysis research notes, comments, and replies. Citation information could not be located for 4 articles. For 45 articles, information about the rounds of review was available, but reviewers’ recommendations were missing.
9. ASR would generally send a revise-and-resubmit manuscript to the original reviewers and one new reviewer. See Bakanic (1986) for details.
10. Because the Web of Science has poor coverage of the social sciences relative to other prominent databases, it may fail to index manuscripts citing the ASR articles under consideration here (Bormann et al. forthcoming). This consideration should not affect the comparison of Web of Science citation counts with each other, as all counts are disadvantaged equally.
11. For the complete list of topics and coding instructions, see Bakanic (1986).
12. Pairwise correlation coefficients between the variables are reported in Appendix Table A3.
13. Kernel density estimates plots, like histograms, visualize distributions (“Kernel Density Estimation” n.d.).
14. We also estimated logistic regressions for the relationship between average review score and whether an article’s citations fell (1) or did not fall (0) in the top 5 percent or 10 percent. In both regressions, the average review score was not associated with the odds of the article representing a citation “hit.” Results are available on request.
15. Statistical tests comparing the mean citations (or mean logged citations) of the two groups at each year lead to the same conclusion and are available from the authors on request.
16. For example, Bornmann et al. (2012) studied citations of manuscripts published by a chemistry journal and found that citations, when controlling for peer review score, were predicted by the chemical subfield.

Acknowledgments

We thank the participants of the Social Theory and Evidence Workshop (University of Chicago) and the Skat25 Conference (Chicago) for helpful comments on previous versions of this research. We also thank the anonymous reviewers, Ben Merriman, and Clark McPhail for their input. All errors are our own.

Notes

1. Prominent recent examples include Li and Agha (2015) and Siler, Lee, and Bero (2015).
2. Reviewers may also fail to accurately assess these other dimensions. Our data do not allow us to address this possibility.
3. For recent reviews, see Bornmann (2011), Lamont (2009), and Lee et al. (2013).
4. Research on revisions is limited, and the impact of revisions on citation is unknown. Goodman et al. (1994) found that revisions reduce “spin.” Decrease in spin may be expected to decrease citations. Two studies of sociology journals found that revisions alter the framing of results (Strang and Siler 2015; Teplitskiy 2015), but the implications of change of frame on citations are unclear.
5. Bornmann and Daniel (2008) came closest to properly taking publishing journals’ Impact Factors into account by comparing published and rejected-and-published articles’ citations with a baseline of all journals within the discipline of chemistry or one of its subfields. Unfortunately, this approach does not preclude the possibility that journals publishing initially rejected manuscripts are still of higher Impact Factors than the field’s average.
6. It does not appear that the authors accounted for the possibility that desk-rejected articles were resubmitted to lower impact journals than articles rejected after review.
7. Five percent of incoming manuscripts were desk-rejected or withdrawn by the author(s). Of the remaining manuscripts, 63 percent were rejected after review, 25 percent were requested to be revised and resubmitted (and 19 percent eventually published), and 9 percent were unconditionally accepted (Bakanic et al. 1987:634).
8. This set includes nearly all of the articles published between 1978 and 1981 and some published in 1982. We exclude from analysis research notes, comments, and replies. Citation information could not be located for 4 articles. For 45 articles, information about the rounds of review was available, but reviewers’ recommendations were missing.
9. ASR would generally send a revise-and-resubmit manuscript to the original reviewers and one new reviewer. See Bakanic (1986) for details.
10. Because the Web of Science has poor coverage of the social sciences relative to other prominent databases, it may fail to index manuscripts citing the ASR articles under consideration here (Bormann et al. forthcoming). This consideration should not affect the comparison of Web of Science citation counts with each other, as all counts are disadvantaged equally.
11. For the complete list of topics and coding instructions, see Bakanic (1986).
12. Pairwise correlation coefficients between the variables are reported in Appendix Table A3.
13. Kernel density estimates plots, like histograms, visualize distributions (“Kernel Density Estimation” n.d.).
14. We also estimated logistic regressions for the relationship between average review score and whether an article’s citations fell (1) or did not fall (0) in the top 5 percent or 10 percent. In both regressions, the average review score was not associated with the odds of the article representing a citation “hit.” Results are available on request.

References

Bakanic, Eunice Yvonne. 1986. “Tracing Social Science: The Manuscript Review Process.” PhD dissertation, University of Illinois at Urbana-Champaign.
Bakanic, Von, Clark McPhail, and Rita J. Simon. 1987. “The Manuscript Review and Decision-making Process.” American Sociological Review 52(5):631–42. doi:10.2307/2095599.
Bakanic, Von, Clark McPhail, and Rita J. Simon. 1989. “Mixed Messages: Referees’ Comments on the Manuscripts They Review.” Sociological Quarterly 30(4):639–54.
Bormann, Lutz. 2011. “Scientific Peer Review.” Annual Review of Information Science and Technology 45:197–245. doi:10.1002/aris.2011.1440450112.
Bormann, Lutz, and Hans-Dieter Daniel. 2008. “What Do Citation Counts Measure? A Review of Studies on Citing Behavior.” Journal of Documentation 64(1):45–80. doi:10.1108/00220410810844150.
Bormann, Lutz, and Werner Marx. 2014. “The Wisdom of Citing Scientists.” Journal of the Association for Information Science and Technology 65(6):1288–92. doi:10.1002/asi.23100.
Bornmann, Lutz, Rüdiger Mutz, and Hans-Dieter Daniel. 2010. “A Reliability-generalization Study of Journal Peer Reviews: A Multilevel Meta-analysis of Inter-rater Reliability and Its Determinants.” *PloS ONE* 5(12):e14331. doi:10.1371/journal.pone.0014331.

Bornmann, Lutz, Irina Nast, and Hans-Dieter Daniel. 2008. “Do Editors and Referees Look for Signs of Scientific Misconduct When Reviewing Manuscripts? A Quantitative Content Analysis of Studies That Examined Review Criteria and Reasons for Accepting and Rejecting Manuscripts for Publication.” *Scientometrics* 77(3):415–32. doi:10.1007/s11192-007-1950-2.

Bornmann, Lutz, Hermann Schier, Werner Marx, and Hans-Dieter Daniel. 2012. “What Factors Determine Citation Counts of Publications in Chemistry Besides Their Quality?” *Journal of Informetrics* 6(1):11–18. doi:10.1016/j.joi.2011.08.004.

Bornmann, Lutz, Andreas Thor, Werner Marx, and Hermann Schier. Forthcoming. “The Application of Bibliometrics to Research Evaluation in the Humanities and Social Sciences: An Exploratory Study Using Normalized Google Scholar Data for the Publications of a Research Institute.” *Journal of the Association for Information Science and Technology*.

Bourdieu, Pierre. 1975. “The Specificity of the Scientific Field and the Social Conditions of the Progress of Reason.” *Social Science Information* 14(6):19–47. doi:10.1177/053901847501400602.

Brunsm, David, Monica Prasad, and Ezra Zuckerman. 2013. “Strategies for Reviewing Manuscripts.” Retrieved March 8, 2016 (http://www.asanet.org/documents/asa/pdfs/Review_Times_in_Sociology.pdf).

Calcagno, V., E. Demoinet, K. Gollner, L. Guidi, D. Ruths, and C. e Mazancourt. 2012. “Flows of Research Manuscripts among Scientific Journals Reveal Hidden Submission Patterns.” *Science* 338(6110):1065–69. doi:10.1126/science.1227833.

Chase, Janet M. 1970. “Normative Criteria for Scientific Publication.” *American Sociologist* 5(3):262–65.

Cicchetti, Domenic V. 1991. “The Reliability of Peer Review for Manuscript and Grant Submissions: A Cross-disciplinary Investigation.” *Behavioral and Brain Sciences* 14(1):119–135. doi:10.1017/S0140525X00065675.

Clemens, Elisabeth S., Walter W. Powell, Kris McIlwaine, and Dina Okamoto. 1995. “Careers in Print: Books, Journals, and Scholarly Reputations.” *American Journal of Sociology* 101(2):433–94.

Cohen, Lawrence E., and Marcus Felson. 1979. “Social Change and Crime Rate Trends: A Routine Activity Approach.” *American Sociological Review* 44(4):588–608. doi:10.2307/2094589.

Cole, S., J. R. Cole, and G. A. Simon. 1981. “Chance and Consensus in Peer Review.” *Science* 214(4523):881–86. doi:10.2307/1686309.

Danthi, Narasimhan, Colin O. Wu, Peibei Shi, and Michael Lauer. 2014. “Percentile Ranking and Citation Impact of a Large Cohort of NHLBI-funded Cardiovascular R01 Grants.” *Circulation Research* 114(4):600–606. doi:10.1161/CIRCRESAHA.114.302656.

Espeland, Wendy Nelson, and Michael Sauder. 2007. “Rankings and Reactivity: How Public Measures Recreate Social Worlds.” *American Journal of Sociology* 113(1):1–40. doi:10.1086/517897.

Espeland, Wendy Nelson, and Mitchell L. Stevens. 2008. “A Sociology of Quantification.” *European Journal of Sociology/Archives Européennes de Sociologie* 49(3):401–36. doi:10.1017/S0003975509000130.

Goodman, Steven N., Jesse Berlin, Suzanne W. Fletcher, and Robert H. Fletcher. 1994. “Manuscript Quality before and after Peer Review and Editing at *Annals of Internal Medicine*.” *Annals of Internal Medicine* 121(1):11–21. doi:10.7326/0003-4819-121-1-19940701-00003.

Guettzkow, Joshua, Michèle Lamont, and Grégoire Mallard. 2004. “What Is Originality in the Humanities and the Social Sciences?” *American Sociological Review* 69(2):190–212.

“Kernel Density Estimation.” N.d. Retrieved October 22, 2015 (https://en.wikipedia.org/wiki/Kernel_density_estimation).

Lakhani, Karim R., and Robert G. Wolf. 2003. “Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects” (SSRN Scholarly Paper No. ID 443040). Rochester, NY: Social Science Research Network. Retrieved March 8, 2016 (http://papers.ssrn.com/abstract=443040).

Lakhani, Karim R., and Eric von Hippel. 2003. “How Open Source Software Works: ‘Free’ User-to-user Assistance.” *Research Policy* 32(6):923–43. doi:10.1016/S0048-7333(02)00095-1.

Lamont, Michèle. 2009. “How Professors Think: Inside the Curious World of Academic Judgment. Cambridge, MA: Harvard University Press.

Leahey, Erin. 2008. “Methodological Memes and Mores: Toward a Sociology of Social Research.” *Annual Review of Sociology* 34:33–53. doi:10.1146/annurev.soc.34.040507.134731.

Leahey, Erin, and James Moody. 2014. “Sociological Innovation through Subfield Integration.” *Social Currents* 1(3):228–56. doi:10.11177/2329496514540131.

Lee, Carole J. 2015. “Commensuration Bias in Peer Review.” *Philosophy of Science* 82(5):1272–83.

Lee, Carole J., Cassidy R. Sugimoto, Guo Zhang, and Blaise Cronin. 2013. “Bias in Peer Review.” *Journal of the American Society for Information Science and Technology* 64(1):2–17. doi:10.1002/asi.22784.

Leimu, Roosa, and Julia Koricheva. 2005. “What Determines the Citation Frequency of Ecological Papers?” *Trends in Ecology & Evolution* 20(1):28–32. doi:10.1016/j.tree.2004.10.010.

Li, Danielle, and Leila Agha. 2015. “Big Names or Big Ideas: Do Peer-review Panels Select the Best Science Proposals?” *Science* 348(6233):434–38. doi:10.1126/science.aaa0185.

Long, J. Scott. 1978. “Productivity and Academic Position in the Scientific Career.” *American Sociological Review* 43(6):889–908. doi:10.2307/2094628.

Merton, Robert K., and Harriet Zuckerman. 1971. “Patterns of Evaluation in Science: Institutionalisation, Structure and Functions of the Referee System.” *Minerva* 9(1):66–100.

Mervis, Jeffrey. 2014. “Peering into Peer Review.” *Science* 343(6171):596–98. doi:10.1126/science.343.6171.596.

Mervis, Jeffrey. 2015. “NIH’s Peer Review Stands Up to Scrutiny.” *Science* 348(6233):384. doi:10.1126/science.348.6233.384.

Mingers, John, and Loet Leydeseidoff. 2015. “A Review of Theory and Practice in Scientometrics.” *European Journal of Operational Research* 246(1):1–19. doi:10.1016/j.ejor.2015.04.002.

Newman, M.E.J. 2009. “The First-mover Advantage in Scientific Publication.” *EPL (Europhysics Letters)* 86(6):68001. doi:10.1209/0295-5075/86/68001.
Nicolaisen, Jeppe. 2007. “Citation Analysis.” Annual Review of Information Science and Technology 41:609–41. doi:10.1002/aris.2007.1440410120

Park, Hyunwoo, Jeongsik (Jay) Lee, and Byung-Cheol Kim. 2015. “Project Selection in NIH: A Natural Experiment from ARRA.” Research Policy 44(6):1145–59. doi:10.1016/j.respol.2015.03.004.

Siler, Kyle, Kirby K. Lee, and Lisa Bero. 2015. “Measuring the Effectiveness of Scientific Gatekeeping.” Proceedings of the National Academy of Sciences 112(2):360–65. doi:10.1073/pnas.1418218112.

Simon, Rita J., Von Bakanic, and Clark McPhail. 1986. “Who Complains to Journal Editors and What Happens.” Sociological Inquiry 56(2):259–71. doi:10.1111/j.1475-682X.1986.tb00087.x.

Smelser, Neil J. 2015. “Sources of Unity and Disunity in Sociology.” American Sociologist 46(3):303–12. doi:10.1007/s12108-015-9260-2.

Strang, David, and Kyle Siler. 2015. “Revising as Reframing Original Submissions versus Published Papers in Administrative Science Quarterly, 2005 to 2009.” Sociological Theory 33(1):71–96. doi:10.1177/0735275115572152.

Teplitskiy, Misha. 2015. “Frame Search and Re-search: How Quantitative Sociological Articles Change During Peer Review.” American Sociologist. Retrieved March 8, 2016 (http://doi.org/10.1007/s12108-015-9288-3).

van den Besselaar, Peter, and Ulf Sandström. 2015. “Early Career Grants, Performance, and Careers: A Study on Predictive Validity of Grant Decisions.” Journal of Infometrics 9(4):826–38. doi:10.1016/j.joi.2015.07.011.

Van Noorden, Richard, Brendan Maher, and Regina Nuzzo. 2014. “The Top 100 Papers.” Nature 514(7524):550–53. doi:10.1038/514550a.

van Raan, A.F.J. 1996. “Advanced Bibliometric Methods as Quantitative Core of Peer Review Based Evaluation and Foresight Exercises.” Scientometrics 36(3):397–420. doi:10.1007/BF02129602.

Waltman, Ludo. 2015. “A Review of the Literature on Citation Impact Indicators.” Retrieved March 8, 2016 (http://arxiv.org/abs/1507.02099).

Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi. 2007. “The Increasing Dominance of Teams in Production of Knowledge.” Science 316(5827):1036–39. doi:10.1126/science.1136099.

Author Biographies

Misha Teplitskiy is a PhD candidate in sociology at the University of Chicago and a graduate research fellow at KnowledgeLab. His research interests include the sociology of science and knowledge, valuation, and computational methods. He is currently investigating biases in peer review decisions using the review files of thousands of manuscripts submitted to the journal PLoS ONE.

Von Bakanic is an associate professor at the College of Charleston. Her research interest include the sociology of science and knowledge, women’s studies, and the study of prejudices. She is currently working on a content analysis of the newspaper articles written about the 1969 Charleston hospital workers’ strike and an analysis of the intake data at a rape crisis organization.