

RESEARCH ARTICLE

Perception of the importance of chemistry research papers and comparison to citation rates

Rachel Borchardt^{1*}, Cullen Moran¹, Stuart Cantrill², Chemjobber³, See Arr Oh⁴, Matthew R. Hartings^{1*}

1 American University, NW, Washington, DC, United States of America, **2** Nature Chemistry, SpringerNature, London, United Kingdom, **3** Chemjobber, Shell, WV, United States of America, **4** Just Like Cooking, Krypton, KY, United States of America

* borchard@american.edu (RB); hartings@american.edu (MH)



OPEN ACCESS

Citation: Borchardt R, Moran C, Cantrill S, Chemjobber , Oh SA, Hartings MR (2018) Perception of the importance of chemistry research papers and comparison to citation rates. PLoS ONE 13(3): e0194903. <https://doi.org/10.1371/journal.pone.0194903>

Editor: Cassidy Rose Sugimoto, Indiana University Bloomington, UNITED STATES

Received: July 8, 2017

Accepted: February 21, 2018

Published: March 28, 2018

Copyright: © 2018 Borchardt et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The dataset file is available from the figshare repository database (DOI [10.6084/m9.figshare.5187529](https://doi.org/10.6084/m9.figshare.5187529)).

Funding: The authors received no specific funding for this work.

Competing interests: Matthew Hartings - Professor Hartings is an Advisory Board member for Chemical and Engineering News, which is published by the American Chemical Society (ACS). The ACS also publishes the Journal of the American Chemical Society, an issue of which is

Abstract

Chemistry researchers are frequently evaluated on the perceived significance of their work with the citation count as the most commonly-used metric for gauging this property. Recent studies have called for a broader evaluation of significance that includes more nuanced bibliometrics as well as altmetrics to more completely evaluate scientific research. To better understand the relationship between metrics and peer judgements of significance in chemistry, we have conducted a survey of chemists to investigate their perceptions of previously published research. Focusing on a specific issue of the *Journal of the American Chemical Society* published in 2003, respondents were asked to select which articles they thought best matched importance and significance given several contexts: highest number of citations, most significant (subjectively defined), most likely to share among chemists, and most likely to share with a broader audience. The answers to the survey can be summed up in several observations. The ability of respondents to predict the citation counts of established research is markedly lower than the ability of those counts to be predicted by the h-index of the corresponding author of each article. This observation is conserved even when only considering responses from chemists whose expertise falls within the subdiscipline that best describes the work performed in an article. Respondents view both cited papers and significant papers differently than papers that should be shared with chemists. We conclude from our results that peer judgements of importance and significance differ from metrics-based measurements, and that chemists should work with bibliometricians to develop metrics that better capture the nuance of opinions on the importance of a given piece of research.

Introduction

Significance, importance and impact are all goals that the chemistry community strives to maximize—a breakthrough discovery, an important finding, a methodology that fundamentally changes the way chemistry research is performed. While individual researchers' goals

analyzed in the submitted manuscript. See Arr Oh - Dr. SAO is an Advisory Board member for Chemical and Engineering News, which is published by the American Chemical Society (ACS). The ACS also publishes the Journal of the American Chemical Society, an issue of which is analyzed in the submitted manuscript. Dr. SAO has been paid by the ACS for articles written for Chemical and Engineering News. Chemjobber - Dr. Chemjobber has been paid by the ACS for articles written for Chemical and Engineering News. Stuart Cantrill - Dr. Cantrill is the editor-in-chief of Nature Chemistry, a peer journal to the Journal of the American Chemical Society. Rachel Borchardt - Ms. Borchardt is the chair of the ACRL Impactful Scholarship and Metrics Task Force, which is cited in the submitted manuscript.

may vary greatly, many may aim to not only transform how chemistry research is performed, but to influence broader areas that transcend academic borders [1]. However, these abstract notions of significance, importance, and impact, particularly beyond the scholarly arena, are difficult to define, and even more difficult to quantify.

In recent years, many funding agencies have included requirements of efforts that move research outside of the laboratory. Among these is the U.S. National Science Foundation (NSF), which has recognized the importance of these goals with their “Broader Impacts” requirements in funding applications. The NSF gives guidance on these areas of impact, which include “Innovating for our future”, “Improving our society”, and “Engaging a wider audience” [2]. Efforts in these areas are meant to benefit society at large along with ensuring that chemistry (and other scientific disciplines) continues to receive popular support and maintains an ability to influence policy decisions.

A chemist’s work, then, includes the efforts expended in the laboratory on a research program as well as the efforts outside of experimentation for the benefit of society. Both types of work are necessary to further the field of chemistry. Therefore, it follows that chemists should be properly evaluated for the totality of their work. Some of the highest-stake evaluation points will occur early on during the career of an independent scientist: applying for tenure and seeking grant funding for new projects. Each of these evaluations is, essentially, a snapshot of a researcher’s research career, but can have very different aims. A tenure or promotion evaluation is generally reviewing a researcher for, among other criteria, impact and excellence in scholarship, while a grant funder is more interested in the potential of their research to have future impact. But what criteria are used to make these determinations about impact, excellence, and potential?

Relatively little insight is generally given into these evaluative processes, but it is well known within the chemistry community that citation-based metrics, most notably citation counts, impact factor, and h-index, play a large role in these evaluations [3,4]. Unfortunately, citation-based metrics may take some time before providing an estimation of the impact of a research program or an individual researcher. There is, therefore, a need to account for immediacy during these evaluations, particularly with grant funding.

Previous studies have found that grant decisions can be predicted, to some extent, by evaluation of h-index and similar metrics [5–7]. A recent study has called into question whether grant funding now serves as a “reward” for scientists who have accrued sufficient citation counts rather than as an award for promising research [8]. Another study found that h-index did not predict future success for junior level faculty [9].

Another area where perceptions of research significance plays a role is in the evaluation of individual manuscripts by journal editors and their referees. These decisions certainly reach beyond determining the potential of an article to generate citations. Editors evaluate significance and potential interest of an article along with how that manuscript fits into the scope of the journal. Referees assess the quality the experiments in an article, give their estimation on how well these experiments fit within the current research climate, and make general evaluations as to whether the importance of the research merits publication within the journal.

It is clear that traditional, citation-based metrics alone cannot be used to fully evaluate research or a researcher, but immediacy must be taken into consideration as part of these evaluations. It is also clear that the efforts of chemists to make their work relevant to society and achieve the goal of significance, as judged by their peers, have not been widely considered or measured.

Radicchi, Weissman, and Bollen recently looked at researchers’ perceptions of individual papers. They found that there was a low correlation between citation count and researchers’ perception of impact outside of instances where respondents were evaluating their own work

[10]. The authors conclude that researchers are only capable of accurately measuring impact of studies in which they were involved. We, however, would postulate that their findings are an indication that citation counts may not be accurately measuring researchers' abstract perception of impact within their field, but that this more abstract view of impact is instead replaced with citation counts when considering their own research. This takes into account that a researcher will likely be familiar with their own metrics, and have been evaluated on them in the past, while for other papers, they are more free to substitute their own definition of impact. Regardless, this disconnect between peer judgements of impact and citations for individual articles seems worthy of further exploration.

Accordingly, in this paper, we examine the broad concepts of significance and impact in order to better understand chemistry researchers' perceptions of individual research articles. We evaluate how individual judgements, based upon different types of significance, relate to commonly-used citation-based metrics as well as other metrics. By doing so we show that researchers do have nuanced and measurable concepts of importance and significance in an attempt to understand how research and researchers might be evaluated when citation counts are not available or relevant. In effect, researchers are able to see different types of significance in an individual research article. We conclude by offering suggestions for how professional chemistry societies can work with bibliometricians to generate guidance for employing metrics that capture the entirety of a researcher's efforts.

Methodology

With the goal of testing the ability of chemists to evaluate research importance in mind, we chose an issue from the *Journal of the American Chemical Society* (volume: 125, issue: 31), published in 2003, 10 years prior to the survey date. This particular journal and issue was chosen to expose survey participants to articles in a wide variety of chemistry subdisciplines and topics from a flagship journal while controlling for journal-level impact, and to have allowed sufficient time for the articles to accumulate citations. We realize that there are a number of journals from which we could have chosen. Assisting in our decision to choose an issue from *JACS*, was that we all really liked the title, '*The JACS Challenge*' for our survey. A survey instrument was developed and tested via personal correspondence with 25 chemists. Upon analysis of this initial data, noting the inability of professional chemists to correctly choose the 3 top cited articles, we decided to expand our efforts to include more respondents. This survey was then replicated via SurveyMonkey.

This study was exempted, after review, by American University's Institutional Review Board (Protocol #13167).

Survey

The survey consisted of 8 questions, including basic demographic information such as research subdiscipline(s) and age. The following four questions were then applied to the same 63 articles from the *JACS* issue:

- Which three papers in the issue do you think are the most 'significant' (your own definition of 'significant' is what is important here)?
- Without looking up numbers, which three papers do you think will have been cited the most to-date?
- Which three papers would you most want to point out to other chemists?

- Which three papers would you want to shout about from the rooftops (i.e., tell anybody about, not just chemists)?

Data

We recruited respondents using a number of blogposts and various solicitations on social media, including the blog 'In the Pipeline'. In total, 408 people responded to the survey over a 2-week period. 45 sets of responses were removed for the following reasons: failing to give or denying consent on the IRB protocol form to use their answers for evaluation and submitting more than the requested number of responses (3) to a question. Removal of participants resulted in 363 participant responses included in our calculations.

Article citation counts for each article were recorded in November of 2013 and November of 2016 from Web of Science. H-index metrics of all corresponding authors were recorded in January of 2016 from Web of Science, and the highest h-index for each paper was used. Page Views were recorded from the journal's website in November of 2016. Mendeley metrics were recorded from Mendeley in November of 2016.

The data set, available online via Figshare, was analyzed using Microsoft Excel [11].

Methodology limitations

Our survey design and execution contain three main limitations. One, the participants in the survey were self-selected from those that are active within online chemistry communication channels, and are only a small subset of the entire field of chemistry. As such, they are not necessarily an accurate representation of the entire field, but their basic demographics are noted below. The second limitation is in the design of the survey itself—namely, that the four central questions were not order randomized. As a result, it is entirely possible that one question influenced the next. As shown in the results, most respondents chose different articles for the four questions, but the possible effects of this influence are unknown and unaccounted for. A third limitation is the open, rather than blind, manner in which the subdisciplines for each article were assigned through an evaluation by the authors of this paper.

We stress to the reader that this survey and study are designed to serve as preliminary analyses designed to illustrate and quantify abstract ideas in a new and interesting way. We invite bibliometricians and other researchers to more deeply delve into our data to examine additional trends not included in our study, including how self-citations may affect differences in total citations and perceived importance.

Results and discussion

[Table 1](#) lists the self-reported demographics of the respondents, who were able to select multiple subdisciplines to accurately and fully describe their training and current duties. The demographics of our respondents were weighted more heavily towards organic and medicinal chemistry than the general membership of the American Chemical Society [12]. We estimate that this is due to the readership of Derek Lowe's blog 'In The Pipeline', which focuses on pharmaceutical chemistry and was likely the largest driver for publicizing this survey [13]. Nevertheless, we were able to attract respondents whose self-associated subdisciplines come from across the full spectrum of chemistry.

Each respondent was asked to choose 3 papers for each question. In looking at the papers that were selected, the highest correlation between the sets of answers was found for papers that were deemed significant and those that were thought to be the highest cited. Correlations among these categories are shown in [Table 2](#). The correlation between papers deemed

Table 1. Respondent demographics.

N = 363	Respondent Demographics								
	Analytical	Biochemistry	Inorganic	Organic	Medicinal	Physical	Polymers	Other	Decline
Subdiscipline									
Respondents	14	52	44	206	60	28	21	15	2
Papers Classified by Subdiscipline (63)	1	14	33	37	3	26	12	19	
Employment	Graduate Student	Postdoc	Faculty Member	Professional within Chemical Industry	Trained Chemist in a Non-Research Field	Decline			
Respondents	93	64	55	117	28	6			
Age	20–30	30–40	40–50	50–60	60–70	Above 70	Decline		
Respondents	140	141	54	23	2	0	3		

<https://doi.org/10.1371/journal.pone.0194903.t001>

significant and those that should be shared with other chemists was higher than the correlation between those that were thought to be the most cited and those that should be shared with other chemists. This result seems to indicate that while chemists do equate significance with citations, the act of sharing papers with other chemists should also be tracked in more detail.

The relationship between sharing with chemists and citations as well as between sharing with chemists and significance is certainly an interesting one. There may be certain papers which chemists find to be significant that might not be expected to show up in citation counts. Prior research focusing on the correlation between scientists storing and sharing papers (specifically through Mendeley reference libraries or over social media) and citation rate has shown that sharing can be predictive of future citations [14,15]. However, these earlier studies never directly looked at the evaluation of significance.

Because citations are the most common way in which the significance of a piece of research is broadly evaluated, we were interested in how well our respondents' evaluations of citations, significance, and sharing correlated with the actual citations. Fig 1 shows several graphs that display total respondent selections with citations as of November of 2013 (when we closed the survey).

Focusing on the top panel (respondent evaluations of number of citations with actual citations), it is apparent that there are fluctuations in both the peer judgement of citation counts and actual citations. That is bound to be expected as there are definitely variations in citation rate, even within journals. Stuart Cantrill, one of the authors on this paper and chief editor of *Nature Chemistry*, has commented on this phenomenon in the context of that journal's JIF, noting the contribution of a small handful of highly cited papers to their overall impact factor, with the majority of articles receiving fewer citations than the impact factor would predict [16]. This phenomenon also led to a published proposal by several leading scientific publishers for the adoption of citation distributions as a replacement for JIF [17].

We see the same kind of fluctuations in the evaluations of our respondents. What is interesting to us are the differences in what the respondents expect to be well-cited papers versus

Table 2. Correlations between answers given for each question.

N = 363	Cited	Significant	Share: Chemists	Share: All
Cited	1	0.9	0.73	0.7
Significant		1	0.82	0.71
Share: Chemists			1	0.64
Share: All				1

<https://doi.org/10.1371/journal.pone.0194903.t002>

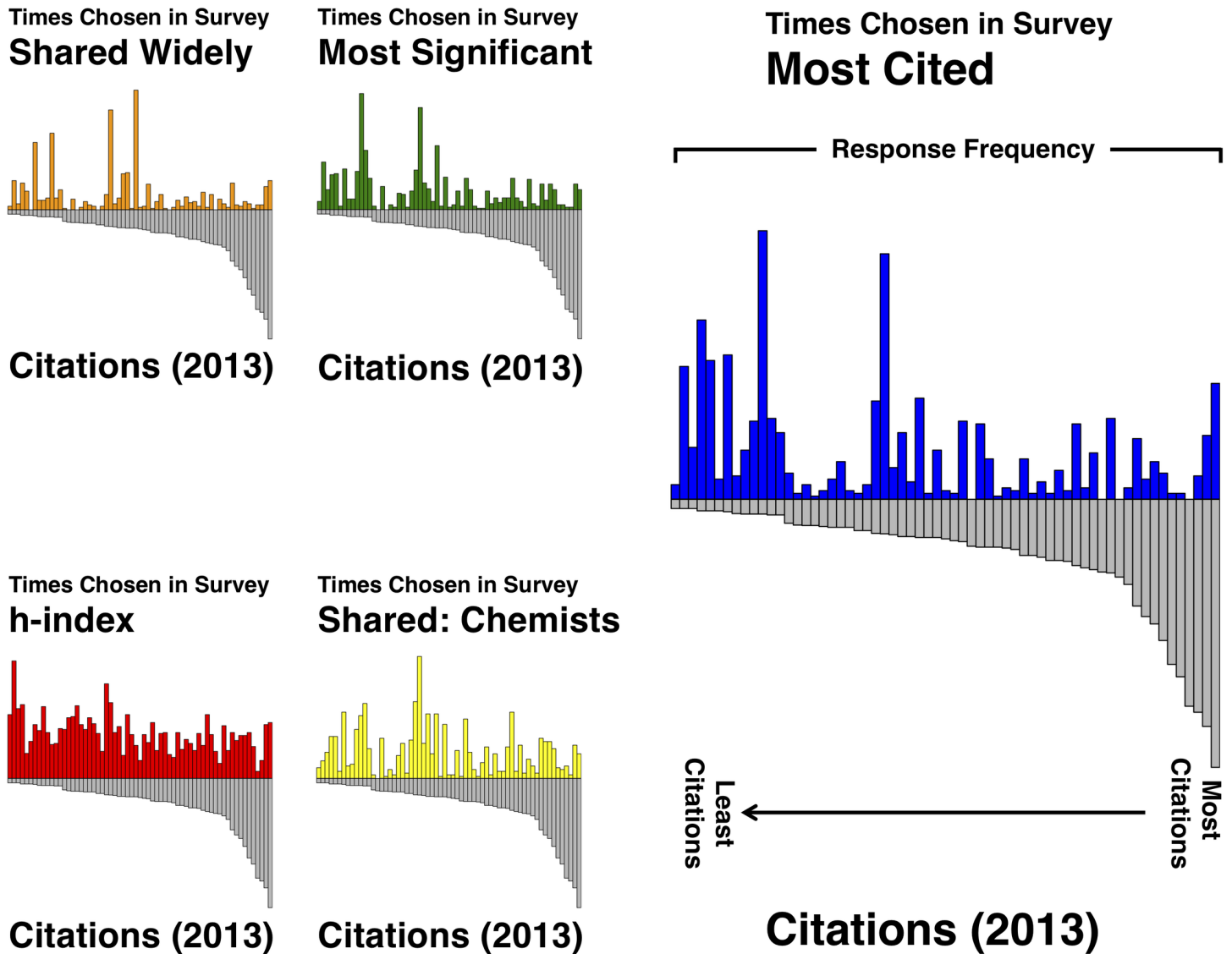


Fig 1. Respondent evaluations and citations (2013) by paper. The top panel shows the composite selections of our respondents for the question asking which papers they thought had the most citations (blue) and the actual number of citations in 2013 (gray). The other panels also include the number of citations (gray) along with: selections for most significant (green, middle left), selections for which should be shared with chemists (yellow, middle right), which should be shared widely (orange, bottom left), and h-index of the corresponding author (red, bottom right) for each of the manuscripts in the journal issue.

<https://doi.org/10.1371/journal.pone.0194903.g001>

which papers are actually well cited. Fig 2 is a more direct way to view the correlation between respondent choices and actual citation counts.

To put these graphs in perspective, we calculated the correlation between the aggregate responses for individual survey questions and the actual citations from both 2013 (the year the survey was taken) and 2016. As shown in Figs 1 and 2, there were a number of papers that were selected disproportionately more than these same papers were actually cited. We were curious if that might be because the respondents were viewing the ‘leading edge’ of chemistry. That is, if these articles were very cutting edge, it might take their citation counts a longer time to build up, like the noted Sleeping Beauty papers [18]. The data doesn’t support this point of view as the correlations, generally, decrease from the 2013 citations to the 2016 citations.

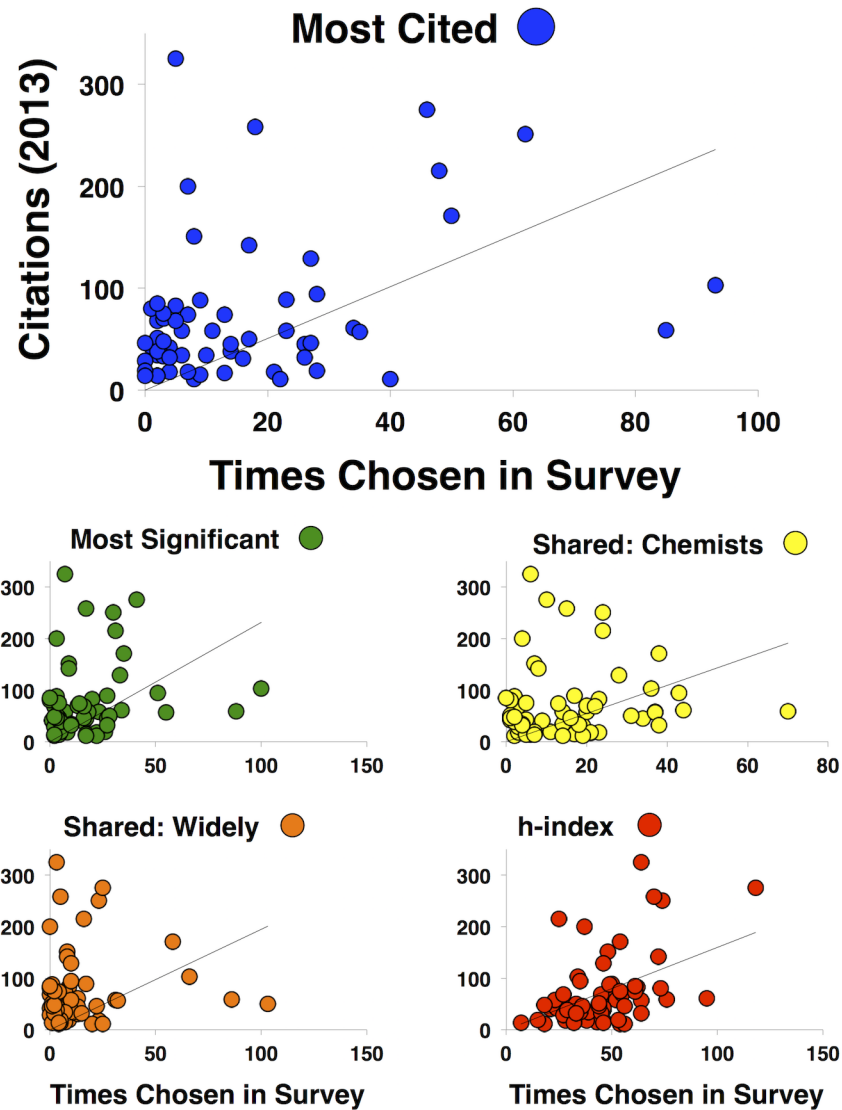


Fig 2. Correlations between respondent choices and citations (2013). The top panel is a graph plotting each papers number of actual citations (2013) versus the number of times it was chosen by our respondents for the question: ‘which three papers do you think have been cited most to-date’ (blue). The other panels also plot actual citations versus survey responses: most significant (green, middle left), shared with chemists (yellow, middle right), and shared widely (orange, bottom left). The number of actual citations versus the h-index of the corresponding author is shown in the bottom right panel (red).

<https://doi.org/10.1371/journal.pone.0194903.g002>

Delving into the individual survey questions, we found generally low correlations between each of the four categories and citations, as well as the correlation with the corresponding author’s h-index, as shown in Table 3. There are several observations that can be made from these comparisons. Primary among these is that corresponding author h-index has a stronger correlation than any of our survey responses to actual citation counts. In some ways, this is unsurprising since both h-index and impact factor play a role in citation count, and h-index is sometimes the strongest predictor of citation count [19,20]. We would caution that h-index and citation count don’t fully capture the notion of significance, as measured by our survey, indicating that, despite their common use in evaluation, true significance cannot be captured

Table 3. Correlation between respondent choices and citation counts from 2013 and 2016.

Correlation with citations (2013)				
Cited	Significant	Share: Chemists	Share: All	h-index
0.33	0.2	0.06	0.10	0.47
Correlation with citations (2016)				
Cited	Significant	Share: Chemists	Share: All	h-index
0.3	0.19	0.06	0.08	0.47

<https://doi.org/10.1371/journal.pone.0194903.t003>

solely by these commonly-used bibliometrics. However, our findings also concur with one study in chemical engineering that shows that author and research group name recognition correlate strongly with citation count, which gives more evidence to the idea that peer judgments are influenced by name recognition [21].

On the subject of true significance, it is worth noting that the survey responses indicate far from perfect agreement. While general trends are certainly demonstrated, the chemists did not all select the same articles for these different proxy measures of abstract importance and significance. However, the fact that there are trends at all show that there is a level of general agreement. Regardless, we posit that human judgement, even limited in scope as presented in these results, is a more accurate reflection of the abstract notion of significance upon which we can then compare metrics such as citation count. We furthermore recognize that the true value of a single article to the academic community as well as society as a whole can probably never be fully represented by any human or metric measures.

The second noteworthy observation from these numbers is that our respondents' judgements of citation numbers were better correlated to actual citations than their estimations of other qualities (significance, shared with chemists, and shared broadly). In some way, this is reassuring; in comparison to the other metrics in the survey, chemists are best at judging citations. This further supports the hypothesis that our respondents, collectively, seem to view citations as being different than significance.

Curious, to us, is the correlation between the articles that would be shared with other chemists and the actual citation count. For the comparison to the citation counts in 2013, the 'shared with chemists' has the lowest correlation of all of the evaluations in our survey. As a citation is literally a way that scientists recommend articles to one another, it is surprising to us that its correlation to actual citation counts isn't slightly higher. One reason for this might be due to the fact that our respondents might equate 'shared with chemists' with a paper being 'interesting' rather than 'useful'. For the sake of this argument, it is instructive to think of two kinds of novelty produced by chemical research: results that lead to a new way of thinking about chemistry and results that lead to a new way of doing chemistry. The latter include reports that are often defined as being 'methods' papers. Our results, when viewed from this perspective may be hinting that there might be recognizable differences between a citable paper and a thought-provoking paper. This conclusion is supported by Li and Thelwall, who observed differences in the prediction strength of F1000 peer judgements for different types of articles, as well as other researchers who note that Mendeley statistics could measure "hidden impact" of some types of research articles [22,23]. While the present study does not fully answer this question, it would be useful to study how chemists view and engage with research that explicitly fall squarely into different categories. These results also hint at the importance altmetrics could play, more generally, in capturing this type of significance that citations alone cannot. Unfortunately, due to the age of the journal issue in question, we were unable to track these articles using altmetrics evaluation tools to confirm whether these papers were, indeed, shared widely through social-media channels.

From our full set of results, we wanted to find out how much the selections made by respondents were dictated by their identification with a specific subdiscipline of chemistry. Before we could do this, we had to determine which subdisciplines each of the papers could accurately be assigned to. Individually, the four authors with graduate training in chemistry (SC, CJ, SAO, and MRH) evaluated each article for their judgments on each article’s subdisciplines. The final decision on subdiscipline was made through a collaborative discussion between these authors. All chemical research relies on influence and techniques that come from multiple subdisciplines. However, we made our decisions based on which subdisciplines most informed a particular study as well as which subdiscipline might be most affected by that study. For an article, any subdiscipline that was suggested by our evaluations was accepted. By our analysis, only 4 articles were described as falling within a single subdiscipline, 37 articles within 2, 21 articles within 3, and a single article designated to 4 subdisciplines.

Table 4 shows correlations that analyze the effects of subdiscipline on respondent choices. First, we looked at how often respondents made selections within their own discipline. Papers that were thought to be highest cited and most shared with chemists were selected from the respondent’s subdiscipline 70% of the time. Papers that were thought to be significant were selected from a respondent’s subdiscipline slightly less often, at 63%.

We also looked to see if selections within subdiscipline correlated better with actual citations than selections outside of subdiscipline. Selections from within subdiscipline and outside of subdiscipline had the same correlation to actual citations in 2013, and both sets had a lower correlation than the aggregate selections. Looking at the paper-by-paper data (Fig 3), there are examples of individual subdisciplines’ ability, or lack thereof, to correlate to actual citations.

In some cases, the respondents’ choices of times cited had a strong correlation to actual citations. Paper 45, ‘A Heterocyclic Peptide Nanotube’ was cited 171 times as of 2013. This paper was selected as being highly cited 50 times, 42 of which coming by respondents that self-identified as having the subdisciplines that characterized that paper. Paper 10, ‘Single-Step *in Situ* Synthesis of Polymer-Grafted Single-Wall Nanotube Composites’ was cited 251 times as of 2013. Our respondents chose this paper as being highly cited 62 times, 34 times from within the subdiscipline and 28 times from outside of the subdiscipline. Paper 5, ‘An Ionic Liquid-Supported Ruthenium Carbene Complex: A Robust and Recyclable Catalyst for Ring-Closing Olefin Metathesis in Ionic Liquids’ was cited 215 times as of 2013. Our respondents chose this paper 48 times, 13 times from within the subdiscipline and 35 times from outside of the subdiscipline.

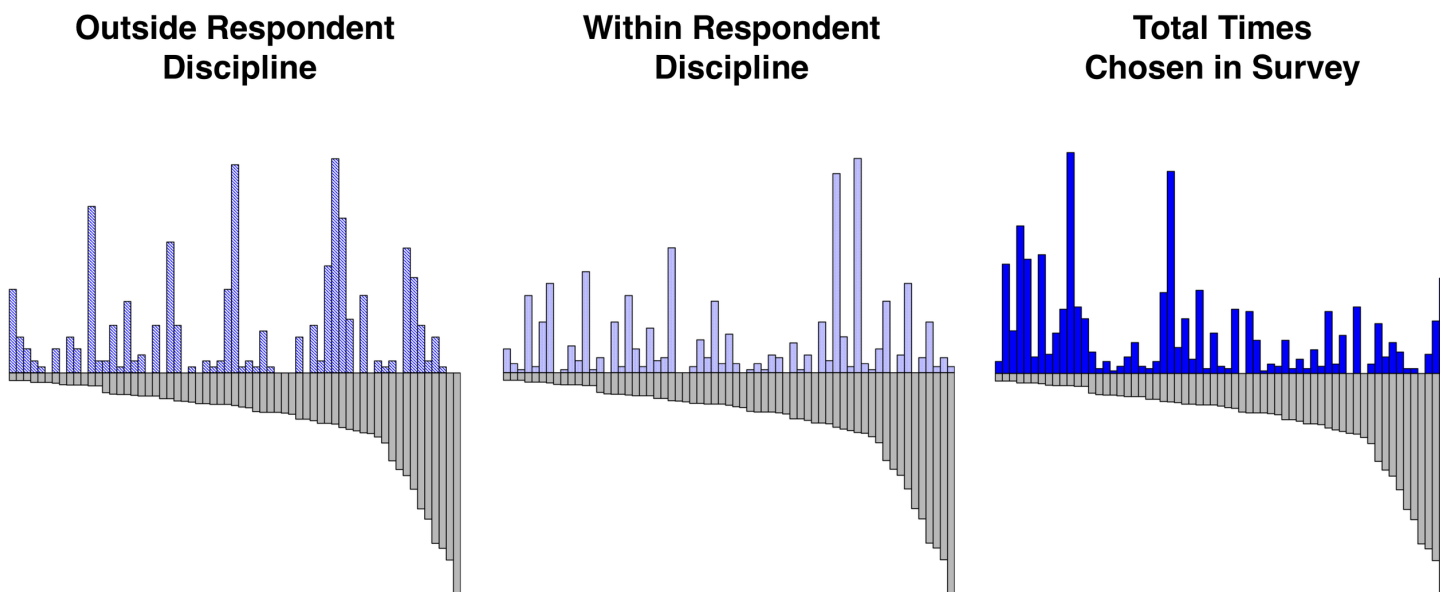
In some cases, our respondents entirely underestimated the citation count of an article. Paper 43, ‘Factors that Determine the Protein Resistance of Oligoether Self-Assembled Monolayers—Internal Hydrophilicity, Terminal Hydrophilicity, and Lateral Packing Density’ was cited 325 times as of 2013. Our respondents chose this paper a total of 5 times, all from within the paper’s subdiscipline.

Table 4. Analysis of selections within and outside of respondents’ subdisciplines.

	Were papers chosen from within respondent’s discipline? Yes. Percentage of articles chosen from respondent’s discipline			
	Cited	Significant	Share: Chemists	Share: All
	69	63	71	51
	Were paper choices made within discipline more accurate than those that came from the respondents with a different discipline? Not Really . Correlation between choice and actual citations			
	Cited	Significant	Share: Chemists	Share: All
In Discipline	0.28	0.18	0.02	0.17
Out of Discipline	0.27	0.12	0.11	-0.01

<https://doi.org/10.1371/journal.pone.0194903.t004>

Response Frequency: Most Cited



Citations (2013)

Fig 3. Respondent selections of most cited by subdiscipline. Each graph shows respondent selection for most cited (left axis, blue) and actual citations as of 2013 (right axis) for each paper. The top panel shows selections from all respondent subdisciplines. The middle panel shows selections where the respondent subdiscipline matches the article subdiscipline. The bottom panel shows selections where the respondent subdiscipline does not match the article subdiscipline.

<https://doi.org/10.1371/journal.pone.0194903.g003>

In some cases, our respondents over-estimated the citation count of an article. Paper 30, ‘Formation of Nanostructured Polymer Filaments in Nanochannels’ was only cited 11 times as of 2013. Our respondents chose this paper 40 times, 36 times from outside of the papers sub-discipline. Paper 18, ‘TNA Synthesis by DNA Polymerases’ was cited 59 times as of 2013. Our respondents chose this paper 85 times, 72 times from within the paper’s subdiscipline.

What this data indicates to us is that human-based prediction and judgement of citation rate is difficult. As seen in other fields looking at the predictive power of F1000, active input from a broadly represented group of chemists is needed to best estimate the citation-impact of an article. But even then, it should be expected that the group will overestimate some papers while underestimating others, as echoed in previous studies [24,25]. Yogi Berra, a baseball player who is noted for his quotability as much as his skill, once said, “It’s tough to make predictions, especially about the future.” Our study shows that it can be just as tough to make predictions about the past.

Carrying forward the ability of diverse perspectives to optimize judgements and predictions, we wanted to see if the aggregate responses to all of our survey questions could better judge citation counts. Table 5 shows the correlations of combinations of responses to 2013 citation counts.

Table 5. Correlation of aggregate survey responses to 2013 citation counts.

Correlation to Citations in 2013			
All Selections: Cited + Significant + Share: Chemists + Share: All	Cited + Significant + Shared: Chemists	Cited + Significant	Cited
0.2	0.22	0.27	0.33

<https://doi.org/10.1371/journal.pone.0194903.t005>

Table 6. Correlations between Mendeley inclusions or page views with citations and survey results.

Correlation to Mendeley inclusions					
2013 Citations	2016 Citations	Most Cited	Most Significant	Share: Chemists	Share: All
0.75	0.76	0.29	0.22	0.17	0.19
Correlation to page views					
2013 Citations	2016 Citations	Most Cited	Most Significant	Share: Chemists	Share: All
0.83	0.85	0.31	0.24	0.20	0.24

<https://doi.org/10.1371/journal.pone.0194903.t006>

What we find from this analysis is that, based upon the questions we asked of our respondents, questions about citation counts were best at actually choosing more highly-cited articles.

We next looked at the correlation between metrics not commonly used in research evaluation and citation counts in relation to our peer judgements. For this study, we have chosen to look at two metrics: inclusion in Mendeley libraries and page views. One reason we chose to look at these metrics is their immediacy—that is, that these metrics can both be gathered and analyzed relatively quickly following an article or even pre-print publication, and thus serve as interesting potential metrics for the evaluation of junior level researchers. Additionally, as mentioned before, they may hold promise in being able to give more nuanced views of impact. Table 6 shows the correlations between these metrics with citation counts and the responses from our survey. As can be seen with the correlations to the 2013 and 2016 citations, both the Mendeley and page-views metric are better at matching relative citation counts than any of our survey responses, reaffirming previous studies that have found correlations between these metrics. These two metrics also show higher correlations to citation rates than does the h-index of the corresponding author. We would like to caution that correlations to these metrics (as we present them) are not predictive and are, likely, formed somewhat in response to the ways these papers are cited. However, what is apparent is that even these metrics fail to replicate what practicing chemists view as significant and important, indicating that, at least for Mendeley, it alone cannot adequately capture these more elusive properties.

Conclusions and outlook

While the analysis of article citations will always be a necessary activity, it is apparent that citation count does not necessarily correlate with perceived importance. Or, put better: that the importance of a research article is only partly captured by its citation rates. This is certainly not a new finding. But we do hope that our study reinforces previous studies and pushes for use of a better metric or set of metrics that captures paper importance that reflects objective analyses (usage of the paper through citation rates and implementation of its findings outside of the research lab) as well as the more subjective point-of-view of active researchers. Returning to the distinction between papers that change the practice of chemistry versus those that change a chemist's point of view, how do researchers engage with these distinct types of research articles? Is there a way to capture nuances of how researchers perceive a research article? Could altmetrics play a role in telling this side of the story? Our results indicate that in order to more closely align current measurements with peer judgement, perhaps the future of chemistry impact evaluation should rely more heavily on perception, peer evaluation, and other qualitative peer judgements than on bibliometrics, in addition to those bibliometrics and altmetrics currently available.

Another aspect that continues to motivate us involves the immediacy of research. A full accounting of the importance of a paper, however importance is defined, cannot be accurately evaluated until long after the research has been published. However, scientists are judged on

the perceived potential of their work before it occurs (while evaluating proposals for funding) and directly after a set of experiments is completed (refereeing manuscripts for journals). The inability of our panel of experts in chemical research to outperform h-index in judging journal citations suggests the views of just a few critical reviewers: journal editors, program officers, article and grant referees, and the scientists performing the work are inadequate for accurately capturing the importance of a research project and result. Perhaps there are other leading indicators, such as engagement with an article on a preprint server, that can better predict and evaluate the importance of a research project prior to publication, provided that a diverse and robust group of peers can be relied on to engage in such activities in an online sphere.

With these findings, we conclude that the time has come for the chemistry community to actively seek better measures for the evaluation of researchers in the context of the broader impacts we describe earlier in the paper. Our suggestion is not a new one, but adds to calls made by others in chemistry and beyond [26–30]. We note that several other disciplines have already investigated and articulated the need for broader impact measures. The most well-defined example of a model for the evaluation of research based on broader impacts is the Becker Model [31]. The Becker Model outlines the impact of research output and activities in five distinct areas of impact: Advancement of Knowledge (which generally corresponds to traditional scholarly impact), Clinical Implementation, Legislation and Policy, Economic Benefit, and Community Benefit. For each area of impact, it includes a list of different research outputs and activities, as well as measures and other indicators of impact. Other examples include practitioners of marketing and policy research, who have suggested ways to track and foster ‘societal impact’, and the development of a task force in academic librarianship, who are charged with creation of a framework for the evaluation of scholarship in a field that is generally considered to be practitioner-driven [32,33].

We specifically call on all international chemical societies to work together to create an evaluative framework for chemistry researchers that go beyond commonly-used bibliometrics to create a framework for the equitable evaluation of research in the context of the broad impacts it provides to scholarship and in the greater society.

Acknowledgments

The authors would like to thank the following chemists for their assistance in the survey trial before its full release: Gavin Armstrong, a contingent of members of the Baran Lab at the Scripps Institute, Paul Bracher, Jessica Breen, Raychelle Burks, Vikki Cantrill, Stuart Conway, Christopher Cramer, Aaron Finke, Katherine Franz, Katherine Haxton, Grant Hill, Ashutosh Jogalekar, Marcell Swart, and Bruce Turnbull.

Author Contributions

Conceptualization: Rachel Borhardt, Stuart Cantrill, Chemjobber, See Arr Oh, Matthew R. Hartings.

Formal analysis: Rachel Borhardt, Cullen Moran, Stuart Cantrill, Chemjobber, See Arr Oh, Matthew R. Hartings.

Project administration: Rachel Borhardt, Matthew R. Hartings.

Supervision: Matthew R. Hartings.

Writing – original draft: Rachel Borhardt, Stuart Cantrill, Matthew R. Hartings.

Writing – review & editing: Rachel Borhardt, Stuart Cantrill, Chemjobber, See Arr Oh, Matthew R. Hartings.

References

- Whitesides GM. Reinventing chemistry. *Angewandte Chemie—International Edition*. 2015; 10.1002/anie.201410884
- National Science Foundation. Broader Impacts [Internet]. [cited 21 Dec 2017]. Available: <https://www.nsf.gov/od/oia/special/broaderimpacts/>
- Stephan P, Veugelers R, Wang J. Reviewers are blinkered by bibliometrics. *Nature*. 2017; 544: 411–412. <https://doi.org/10.1038/544411a> PMID: 28447652
- Cameron BD. Trends in the Usage of ISI Bibliometric Data: Uses, Abuses, and Implications. *portal Libr Acad*. The Johns Hopkins University Press; 2005; 5: 105–125. <https://doi.org/10.1353/pla.2005.0003>
- Hirsch JE. Does the H index have predictive power? *Proc Natl Acad Sci U S A*. National Academy of Sciences; 2007; 104: 19193–8. <https://doi.org/10.1073/pnas.0707962104> PMID: 18040045
- Rezek I, McDonald RJ, Kallmes DF. Is the h-index Predictive of Greater NIH Funding Success Among Academic Radiologists? *Acad Radiol*. 2011; <https://doi.org/10.1016/j.acra.2011.06.017> PMID: 21873082
- Mazloumian A. Predicting Scholars' Scientific Impact. *PLoS One*. 2012; 7. <https://doi.org/10.1371/journal.pone.0049246> PMID: 23185311
- Parra LC, Hirsch L. Award or Reward? Which comes first, NIH funding or research impact? 10.1101/193755
- von Bartheld CS, Houmanfar R, Candido A. Prediction of junior faculty success in biomedical research: comparison of metrics and effects of mentoring programs. *PeerJ*. 2015; 3: e1262. <https://doi.org/10.7717/peerj.1262> PMID: 26421238
- Radicchi F, Weissman A, Bollen J. Quantifying perceived impact of scientific publications. *J Informetr*. Elsevier Ltd; 2017; 11: 704–712. <https://doi.org/10.1016/j.joi.2017.05.010>
- Borchardt R, Hartings M. JACS Challenge Data Set [Internet]. 2017 [cited 7 Jul 2017]. 10.6084/m9.figshare.5187529
- Marchant S, Marchant C. Salaries 2013: Analysis of the American Chemical Society's 2013 Comprehensive Salary and Employment Status Survey. Washington, DC; 2013.
- Lowe D. In the Pipeline [Internet]. [cited 6 Mar 2017]. Available: <http://blogs.sciencemag.org/pipeline/>
- Li X, Thelwall M, Giustini D. Validating online reference managers for scholarly impact measurement. *Scientometrics*. Springer Netherlands; 2012; 91: 461–471. <https://doi.org/10.1007/s11192-011-0580-x>
- Priem J, Piwowar HA, Hemminger BM. Altmetrics in the wild: Using social media to explore scholarly impact [Internet]. 2012 Mar. Available: <http://arxiv.org/abs/1203.4745>
- Cantrill S. Nature Chemistry's 2014 impact factor citation distribution. In: *The Sceptical Chymist*, *Nature Chemistry* [Internet]. 2015 [cited 3 Jun 2017]. Available: <http://blogs.nature.com/thescepticalchymist/2015/12/nature-chemistrys-2014-impact-factor-citation-distribution.html>
- Lariviere V, Kiermer V, MacCallum CJ, McNutt M, Patterson M, Pulverer B, et al. A simple proposal for the publication of journal citation distributions. *bioRxiv*. 2016; Available: <http://biorxiv.org/content/early/2016/09/11/062109>
- Ke Q, Ferrara E, Radicchi F, Flammini A. Defining and identifying Sleeping Beauties in science. *Proc Natl Acad Sci U S A*. National Academy of Sciences; 2015; 112: 7426–31. <https://doi.org/10.1073/pnas.1424329112> PMID: 26015563
- Costas R, Bordons M. The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *J Informetr*. 2007; 1: 193–203. <https://doi.org/10.1016/j.joi.2007.02.001>
- Walters GD. Predicting subsequent citations to articles published in twelve crime-psychology journals: Author impact versus journal impact. *Scientometrics*. Kluwer Academic Publishers; 2006; 69: 499–510. <https://doi.org/10.1007/s11192-006-0166-1>
- Peters HPF, van Raan AFJ. On determinants of citation scores: A case study in chemical engineering. *J Am Soc Inf Sci*. Wiley Subscription Services, Inc., A Wiley Company; 1994; 45: 39–49. [https://doi.org/10.1002/\(SICI\)1097-4571\(199401\)45:1<39::AID-ASI5>3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1097-4571(199401)45:1<39::AID-ASI5>3.0.CO;2-Q)
- Li X, Thelwall M. F1000, Mendeley and traditional bibliometric indicators. *Proc 17th Int Conf*. 2012; Available: https://www.researchgate.net/profile/Mike_Thelwall/publication/236001315_F1000_Mendeley_and_traditional_bibliometric_indicators/links/02e7e5183d3452ff86000000.pdf
- Mohammadi E, Thelwall M, Haustein S, Larivière V. Who reads research articles? An altmetrics analysis of Mendeley user categories. *J Assoc Inf Sci Technol*. John Wiley and Sons Inc.; 2015; 66: 1832–1846.
- Wardle DA. Do “Faculty of 1000” (F1000) ratings of ecological publications serve as reasonable predictors of their future impact? *Ideas Ecol Evol*. 2010; 3: 11–15. <https://doi.org/10.4033/iee.2010.3.3.c>

25. Waltman L, Costas R. F1000 recommendations as a potential new data source for research evaluation: A comparison with citations. *J Assoc.* 2014; Available: <http://onlinelibrary.wiley.com/doi/10.1002/asi.23040/full>
26. Declaration on Research Assessment [Internet]. [cited 1 Jan 2017]. Available: <http://www.ascb.org/dora/>
27. Kenna R, Mryglod O, Berche B. A scientists' view of scientometrics: Not everything that counts can be counted. 2017; <https://doi.org/10.5488/CMP.20.13803>
28. Henning V, Gunn W. Peer review: why each research community should define the metrics that matter to them. *The Guardian*. 6 Sep 2012. Available: <http://www.theguardian.com/higher-education-network/blog/2012/sep/06/mendeley-altmetrics-open-access-publishing>. Accessed 30 Apr 2014.
29. Académie des Sciences, Leopoldina, Royal Society. Statement by three national academies on good practice in the evaluation of researchers and research programmes [Internet]. 2017. Available: <https://royalsociety.org/-/media/policy/Publications/2017/08-12-2017-royal-society-leopoldina-and-academie-des-sciences-call-for-more-support-for-research-evaluators.pdf>
30. Reward research that changes society. *Nature*. 2018; 553: 5–5. <https://doi.org/10.1038/d41586-017-08943-6> PMID: 29300018
31. Bernard Becker Medical Library. Assessing the Impact of Research | Becker Medical Library [Internet]. [cited 16 Feb 2017]. Available: <https://becker.wustl.edu/impact-assessment>
32. Ozanne JL, Davis B, Murray JB, Grier S, Benmecheddal A, Downey H, et al. Assessing the Societal Impact of Research: The Relational Engagement Approach. *J Public Policy Mark.* 2016; 9156: jppm.14.121. <https://doi.org/10.1509/jppm.14.121>
33. Association of College & Research Libraries. Impactful Scholarship and Metrics Task Force | Association of College & Research Libraries (ACRL) [Internet]. 2017 [cited 21 Dec 2017]. Available: <http://www.ala.org/acrl/aboutacrl/directoryofleadership/taskforces/acr-tfism>