

Understanding, Evaluating and Reporting Research Productivity and Impact

Julienne M. Krennrich, Assistant Director of Research Initiatives, Engineering Research Institute, Iowa State University

Arun K. Somani, Associate Dean for Research, College of Engineering, Iowa State University

Martin H. Spalding, Associate Dean for Research, College of Liberal Arts & Sciences, Iowa State University

A simple, holistic definition of achieving excellence in research is to produce sustained, high-impact discoveries and innovations. They can be fundamental insights into particular areas/topics, innovative applications of known technologies or novel solutions to complex problems. The key question for administrators is how best to understand, evaluate and report research productivity and impact.

Historical Impact Measurement

The peer review system has historically played a large role in measuring impact.¹ Unfortunately, the system is far from perfect. Namely ground-breaking (high-impact) research is often years, if not decades, ahead of its time. For most of the typical advances in research, however, the system works well, particularly within fields. As scientific research has matured, the growth and fine-tuning of sub fields and sub, sub fields has made it increasingly difficult to compare impacts across disciplines, not to mention across departments.

The traditional measures of impact are: publications, citations, student and postdoc involvement, funding profile and technology transfer. The h-Index², named for its founder Jorge E. Hirsch, a measure combining publications with citations, was developed as a way of measuring individuals' career achievements, but de-

pending on the completeness of the publication-tracking system, faculty-to-faculty comparisons within the same discipline are difficult to compare. For example, consider Faculty A and Faculty B from the same department, who are roughly the same academic age:

Web of Science

Faculty A: h-index 20, total cites 1607, average 6.56

Faculty B: h-index 29, total cites 4803, average 14.04

Google Scholar (Removed 0-cites and unrecognized sources)

Faculty A: h/g-index 42/70, total 6520, average 23.88

Faculty B: h/g-index 33/71, total 6721, average 19.82

Microsoft Academic Search

Faculty A: h/g-index 29/46, total 3202, average 15.03

Faculty B: h/g-index 18/41, total 2045, average 12.62.

Which person is the better performer? Which source should we believe/trust? Do we believe in “publish or perish” or “publish and change”? Another question is whether a citation implies a positive or negative impact. There is a bias toward reporting only positive impacts and with an additional pressure that more is always better.

ent departments in the same school. Consider departments 4 & 5; the former expended more than twice the amount of dollars, but only supported 1/3 the number of PhD students and 28% more master's students than the latter. Clearly department 5 is much more PhD focused, yet requires significantly fewer research dollars to support those students. Departments 3 & 7 expense about the same level of research dollars, but the latter confers 20-30% more degrees (both PhD and MS). Of the eight departments, which is the most successful? Additionally, what is the

Dept.	Research Expenditure	PhD students enrolled	PhD degrees conferred	MS students enrolled	MS degrees conferred
1	\$410,020	0.8	0.14	0.6	0.29
2	\$297,495	1.1	0.22	1.2	0.62
3	\$518,728	2.6	0.40	0.3	0.27
4	\$393,437	1.2	0.26	2.9	1.29
5	\$182,582	3.6	0.58	2.1	0.82
6	\$83,616	2.0	0.30	1.9	0.82
7	\$597,723	2.5	0.52	1.1	0.40
8	\$483,957	3.4	0.50	3.2	1.30

Table 1: Three-year averaged research expenditures and graduate students supported per tenure track faculty

Over the past 30 years, the Research Enterprise in the United States has seen amazing growth in the competition for research dollars (State, Private and Federal). In many areas, growth in the scientifically trained workforce has continued, but the trend in available research dollars is decidedly negative. Universities have consequently begun to view research funding itself as a measure of impact and productivity. However, the funding required to excel differs significantly across disciplines. See, for example, the data in Table 1; these could be eight identical departments in different schools or eight differ-

ent departments in the same school. Consider departments 4 & 5; the former expended more than twice the amount of dollars, but only supported 1/3 the number of PhD students and 28% more master's students than the latter. Clearly department 5 is much more PhD focused, yet requires significantly fewer research dollars to support those students. Departments 3 & 7 expense about the same level of research dollars, but the latter confers 20-30% more degrees (both PhD and MS). Of the eight departments, which is the most successful? Additionally, what is the

Impact of a single paper

How accurately can we measure the impact of a single paper? A few reasonable papers published in a high impact journal may disproportionately impact the h-index, as others tend to cite those papers more than other more innovative or ground-breaking papers published in not-so-highly-reputed journals. Additionally, significance is traditionally placed on peer-reviewed publications,

but non-peer-reviewed publications can also be highly impactful. In future, as open access journals, which enable much faster publication times by generally employing fewer reviewers for an individual paper, proliferate, more and more people will begin publishing in these journals simply because it takes so long to get papers published in traditional journals. This could be particularly true for cutting edge technology areas. The true impact of these papers might be missed if one only considered “journal impact factor” when ranking publications.

Journal Impact Factor

Although not all journals were created equal, the impact factor is flawed. The number of citations per “eligible” article over time can be misleading as actual distribution of citations is skewed. Journals can also game the system, by being extremely selective about what papers they accept. For example, 89% of Nature's 2004 impact factor was generated by only 25% of the articles.³ Thus, journal-level metrics are inadequate at capturing the significance of individual papers. We believe the traditional model of peer-reviewed journals should and will necessarily change.

Finally, there is a conflict between objectivity and integrity. Words like “positive”, “significant”, “negative” or “null” are common, but are misleading, because all results are equally relevant to science. Meta-analyses have extensively documented an excess of positive and/or statistically significant results in fields and subfields. Confronted with a “negative” result, a scientist might not publish it or may turn it into a positive result. Additionally, quantitative studies have shown

that financial interests can influence the outcome of research. How do we avoid/minimize such factors?

Today

In 2013, the Center for the Study of Interdisciplinarity at the University of North Texas released a list of 56 positive and negative measurements of impact, including many new internet-technology enabled factors or “altmetrics”.⁴ The NSF has recently changed their policy on CVs asking for “research products” rather than just “publications”.⁵ We are looking to create more **impactful** discoveries, not just more. Analytics should be able to help us more accurately measure impact, in part by allowing us to track more outputs as well as more accurately track the traditional ones. This then enables us to compute a more accurate estimate of the return on investment (ROI). We can use analytics to help set investment goals, set expectation goals on ROI, set priorities, decide what factors to consider and understand the qualitative impact of quantitative data.

Impact of a single grant

It is possible to gauge the impact of a single grant by tracking the following:

- Publications enabled by the funding;
- Intellectual property enabled by the funding;
- Student/postdoctoral training enabled by the funding;
- Impacts on the discipline and outside the discipline—e.g., h-index, news articles, etc.

Taken together, these metrics can provide a qualitative measure of the grant, but it may be years before an accurate measure can be made. There is an inherent time lag (see Figure 1) in achieving

outputs after dollars are allocated. Academia is a high time constant system. Typically, it takes a new assistant professor five-six years before s/he is productive. Similarly, the initial preparation of a new PhD student takes two years. Expenditures on a grant generally trail by a year, publications trail by two years and patents trail by many years. Generally, grants and contracts are not tagged with information about timing, # of students and output publications to enable such an analysis. Can we measure the impact of a body of research beyond simple measures of productivity, intellectual property and workforce training?

Multi-investigator grants

As research funding becomes more and more precious, the trend is toward fewer, but larger grants with many, many researchers. Might there be a negative impact associated with them? They are far more complex to run and the leadership teams of these Centers tend to spend a majority of time administering research, rather than actually doing research.* For the institution, funds management becomes more complex as the dollar amount goes up. There are known cases in which a PI had to forfeit a sizable sum of grant dollars because of mismanagement. Other drawbacks include a tendency to fund more successful researchers within an institution leading to a situation in which the “haves” receive even more and the “have nots” continue to struggle.

* High energy physics provides an example of a field that has historically been organized this way out of necessity (and with great success),

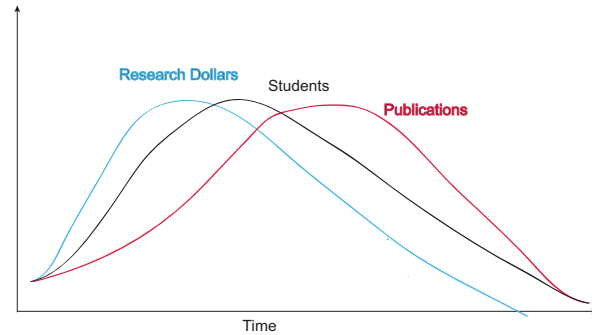


Figure 1: Cartoon of the inherent time lag in academic research

Who wants what data?

Provosts, Deans and Department Chairs all want research output data and are increasingly using fee-based Internet mining services to get it. But the source lies in our institutions: the faculty themselves keep (or can keep) detailed data about each output. We need to build trust so that they will more willingly share these data. We also need to build an output tracking system that is as accurate and automated as possible, so faculty will use it. As one moves up the hierarchy, the level of detail needed in the data decreases. Provosts and Deans are primarily interested in aggregated quantitative data to support qualitative models; the details of individual records are superfluous. So the system should enable increasing aggregation as the data are propagated up the hierarchy.

As a corollary, the inherent anonymity of aggregated data should encourage faculty response and cooperation in sharing. If needed, the stick of tenure and promotion (P&T) portfolio preparation could be used, i.e., require that faculty prepare

but the majority of researchers traditionally have little experience working in large, multi-institutional teams.

these data and include them in their P&T portfolios.

Model: Topical h-Index

Having accepted that not all science/scholarship requires the same amount of money to produce high impact and that sustained funding over a lifetime with many smaller grants will outweigh one giant grant that cannot be sustained, we propose a topic-based evaluation model. Our model seeks to measure researcher output more holistically by grouping publications by researcher-defined topics and computing an equivalent h-index for an entire topic. Since many grants may be used to support a single topic, this would alleviate the time lag in the system by collecting publications on a topic and not just as a result of a single grant. Using appropriate weight factors we would include citations, intellectual property and follow on, such as news articles. This would enable multiple papers with low-medium citations to be weighted more, thereby more accurately measuring a researcher's contribution to a topic over a lifetime.

As an example, we took the Google Scholar citations of the above faculty A and B and, to keep the analysis simple, analyzed only the papers that had at least 20 citations. We categorized papers under topics based on their titles. Here is the distribution:

Faculty A: Number of topics: 30

Total citations: 4846

Distribution:

8 topics with 200 or more citations
20 topics with 100 or more citations
25 topics above 50 or more citations
30 topics above 30 or more citations

Faculty B: Number of topics: 24

Total citations: 6205

Distribution:

7 topics with 200 or more citations
15 topics with 100 or more citations
24 topics with 50 or more citations
24 topics with 30 or more citations

Notice that Faculty A had many papers that had citations between 5 and 20 that were not included in the analysis. Faculty B had the most citations for papers that were covered by the topics.

We consider a contribution to be truly impactful if the topic has been cited 50 or more times. Under this condition, Faculty A is more effective than faculty B, because faculty A has influenced more areas and had a large impact in many of them. On the other hand faculty B had a larger impact on fewer topics (with citations >100 or >200). We noticed that some topics, which had multiple publications (i.e., a conference version followed by a journal version or a follow-on conference paper), turned out to be equally or more impactful than a topic that had only one paper, whether or not that paper contributed to the individuals h-index. Additionally, we must keep in mind that more often than not, a large number of citations are for papers that are tutorial in nature, such as a book. They have a higher education value and therefore attract a higher number of citations in comparison to contributions based on research. This happens to be the case with Faculty B in our example.

How to Collect Data?

To best collect the kind of data needed to do our example analysis on a larger scale, researchers first must register with a publication-tracking service, e.g.,

Google Scholar. Google will automatically find your publications based on your name and institution, but unless you have an uncommon name, there is typically a lot of manual work involved in cleaning up your publication list. One solution to this “identity problem” is to register with ORCID, “an open, non-profit, community-based effort [that] provide[s] a registry of unique researcher identifiers and a transparent method of linking research activities and outputs to these identifiers”.⁶ After researchers obtain their unique ID number, every publication will contain a name and unique number, thereby facilitating the automation of adding publications to a profile. The ORCID approach will also help alleviate the problem of Google treating output that is not published in a peer-reviewed journal as if it were. Only peer-reviewed publications would contain the identifying ORCID number.

Most journals already use topic keywords or topic numbers. These would need to accompany publications in the profile so that the publications can be grouped according to topic. Additionally users should be able to input pointers back to previous papers, providing more definition to the topic. In this way, the system is user driven to increase accuracy, but also enables faculty to engage in the data collection and sharing process. It also allows for more than just “beans” to be counted—users can determine what output (other than publications) is included in a topic. With a little effort to educate faculty on how to use this reporting

format, this is a system that should not be too difficult to maintain. We envision that faculty would annually report such data to their Department Chairs.

Tagging grants with data such as number of graduate students being supported and number of degrees conferred will be more time intensive. It will require some effort from grants administration personnel, but over time, the data produced would be very valuable, so it is worth investing in the effort up front. Generally faculty already list publication outputs and students and personnel supported/trained in technical reports to sponsors, so that tagging is already occurring, just not tracked.

Conclusions

On the whole, we are moving forward; we are beginning to understand how technology and metrics can help us perform better evaluations, but we are still in the experimentation stage.

References:

- 1 Priem, Jason. Beyond the paper. *Nature* 495, 437-440 (2013).
- 2 Hirsch, Jorge E. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Science USA* 102, 16569-16572 (2005).
- 3 Campbell, Philip. Escape from the impact factor. *Nature* 8, 5-7 (2008).
- 4 Brown, Keith Wayne. 56 indicators of impact, <http://cas-csid.cas.unt.edu/?p=4475> (2013).
- 5 Piwowar, Heather. Value all research products. *Nature* 493, 159 (2013).
- 6 ORCID: Connecting Research and Researchers, <http://orcid.org>.