

SIX

Measuring scholarly impact

I do a lot of work with colleges and universities, and study countless catalogs, and it won't surprise you to hear that almost every college catalog in this country still lists teaching, research, and service as the priorities of the professoriate. (Boyer, 1996, p. 22)

In the last speech of his life, Ernest Boyer, then president of the Carnegie Foundation for the Advancement of Teaching, explained the landscape of measuring scholarly impact. He also wanted to change it. Boyer wanted to expand the way academics think about the third leg of the three-legged table of scholarly impact: 'It won't surprise you either that at tenure and promotion time, the harsh truth is that service is hardly mentioned.' Even more disturbing to Boyer was the penalty against faculty who did that kind of work. He observed that researchers who spent time on 'so-called applied projects frequently jeopardize their careers' (Boyer, 1996, p. 22). Speaking in October 1995 on the cusp of the emerging popular Internet, Boyer issued a loud and clear call for a less insular university and for academics to become 'more vigorously engaged in the issues of our day' (Boyer, 1996, p. 28). He would not live to see the rise of networked scholarship – Boyer died just two months after he gave that speech – but in many ways being a scholar in the digital era realizes much of what he envisioned.

Bonnie Stewart, in her recent analysis of how Boyer's vision of scholarship is playing out in the digital era, finds that the networked engagement of digital scholars fulfills what he imagined and takes it even further (Stewart, 2015). Boyer said there was a need to recognize, and value, different kinds of scholarship. Boyer wanted college and university faculty to become 'vigorously engaged with the issues of

our day’ but there would need to be some way to recognize and value different kinds of scholarly work, which he called the ‘scholarship of engagement’ (Boyer, 1990). Under the broad category of the ‘scholarship of engagement,’ Boyer enumerates five types of scholarship: (1) discovering knowledge, (2) integrating knowledge, (3) sharing of knowledge, (4) the application of knowledge, and (5) teaching (Boyer, 1990). In her perceptive analysis, Stewart uses Twitter updates from academics to illustrate each of Boyer’s five types of scholarship. Stewart’s work suggests that digitally networked scholarship embodies Boyer’s initial aim of broadening scholarship itself through fostering extensive cross-disciplinary, public ties and rewarding connection, collaboration, and curation between individual scholars, rather than through their institutions or professional roles. Stewart’s point is clear: the scholarship of engagement has arrived.

Digital technologies have transformed the infrastructure of participation in scholarship in ways that Boyer might never have imagined. The academically focused search engine Google Scholar uses algorithms to search the text of millions of peer-reviewed journal articles, including those behind publishers’ paywalls. A majority of scholars now use it. A recent survey among scientists found that roughly 60% use Google Scholar regularly (van Noorden, 2014).

Anurag Acharya, who co-created and still runs Google Scholar, is an example of someone who is doing the scholarship of engagement that Boyer called for, both applying his knowledge of algorithms to the vast body of academic literature and making it more widely available. Acharya’s work is also having an impact. He recalls:

I came to Google in 2000, as a year off from my academic job at the University of California, Santa Barbara. It was pretty clear that I was unlikely to have a larger impact [in academia] than at Google – making it possible for people everywhere to be able to find information. (van Noorden, 2014)

The other people working on Google Scholar are ‘all, in part, ex-academics’, Acharya explains. He and his colleagues collaborate on one of the digital technologies that are changing the infrastructure of being a scholar. Part of the larger and profitable Internet giant Google, the smaller Scholar operation does not make money. ‘The primary role of Scholar is to give back to the research community, and we are able to do so because it is not very expensive, from Google’s point of view’, explains Acharya (van Noorden, 2014). In supporting the non-

revenue-generating Scholar, Google is playing a role that foundations and universities once played in creating non-for-profit presses.

To be absolutely clear, the non-profit Scholar is made possible precisely because it exists within the larger and very profitable Google, which is essentially a digital media advertising company (Vaidhyanathan, 2011). This should concern scholars for a variety of reasons, including the surveillance and threats to privacy, and the constant threat that services we come to rely on when they are offered at no cost will suddenly only be available for a premium monthly charge (Vaidhyanathan, 2011). Despite these concerns, a survey found that a majority of scholars now use this search service and find it beneficial for their work (van Noorden, 2014).

‘The benefits that Scholar provides ... are very significant’, Acharya contends (van Noorden, 2014). But how would those ‘very significant benefits’ be evaluated in the academy? If Acharya had stayed in his academic position and developed the same innovative search platform there, how would his university measure the impact of his work? It is impossible to know for sure, but in all likelihood Acharya would have been discouraged from working on such a large, collaborative and applied project until after tenure. If he had disregarded such advice and continued working on it, he could have ended up as one of the cases Boyer mentioned of scholars who are punished for doing applied scholarship.

Anurag Acharya’s career and contribution to being a scholar in the digital era raises questions about how measuring scholarly impact is changing. Most of the attention to the ‘priorities of the professoriate’ – teaching, research and service, as Boyer named them – is on scholarship, and on quantitative measures of impact. To appreciate both Boyer’s call for the scholarship of engagement and Acharya’s contribution to the transformations in being a scholar, it is important to understand the influence of Eugene Garfield on the whole ecosystem of measuring scholarly impact.

The (unintended) impact of Eugene Garfield on academia

‘Is there anything comparable to your impact worldwide?’, an interviewer asked Eugene Garfield, comparing his influence to that of Sputnik on space technology (Hargittai, 1999, p. 26). Garfield deflected the interviewer’s comparison as hyperbole, but then wondered on his own, ‘I can’t imagine how you would evaluate the impact of my work.’

How would you measure it?’ (Hargittai, 1999, p. 26). His rhetorical question is deeply ironic, given Garfield’s outsized influence on how academics measure impact (Jasco, 2010).

If you dig very deep at all into how academics think about measuring scholarly impact today, you are bound to encounter the work of Eugene Garfield. A native of the Bronx, Garfield grew up in a Jewish and Italian household, where his early years were shaped by ‘two uncles who were Marxists’ and an absentee father who was a successful businessman (Small, 2007). Garfield would go on to become an influential figure, and a wealthy man, because of the information tools he created. His tools changed the way scholars are employed, professors at universities are given tenure, and research journals are judged for their quality (Hargittai, 1999, p. 26; Jasco, 2010). He created the *Science Citation Index* (SCI), the *Arts and Humanities Citation Index* (AHCI) and the *Social Sciences Citation Index* (SSCI) (Garfield, 1955). For scholars working today, the SCI and SSCI are the standard metrics for assessing scholarly impact of research published in peer-reviewed journals. Divided by disciplines, each index tracks the number of times an article is cited in other journals. Every scholar who has published an article in a peer-reviewed science or social science journal has an entry in the SCI or SSCI. Associated with each name is a count and an indexed list of the citations their work has garnered.

A hybrid scientist and information specialist, Garfield holds degrees in both chemistry and library science, and he earned a PhD in structural linguistics, with a dissertation project that combined chemistry and library science. Born in 1925, he developed his citation analysis in the 1950s. Retired now, but still active, Garfield founded the Institute for Scientific Information (ISI), a company that eventually became one of the world’s largest commercial providers of scientific research data. Garfield is widely credited with founding the fields of bibliometrics (the statistical analysis of written publications) and scientometrics (the study of measuring and analyzing science, technology, and innovation). Garfield has also been called the ‘grandfather of Google Scholar’ for his innovative work on citation indexing, which laid the theoretical basis for the algorithm underneath the popular search platform (Bensman, 2013).

‘Citation analysis exposed the political nature of East European science academies – many academicians were administrators, not world-class scientists’, Garfield explained to an interviewer (Hargittai, 1999, p. 26). Garfield developed citation indexing as a way to spot scientific trends and to trace how an idea flows through the scholarly literature (Garfield, 1955). Garfield extended his work with citation

analysis to the extent that he could predict Nobel laureates with a good degree of accuracy (Garfield and Malin, 1968). Disrupting the political nature of promotion in the academy was an unintended consequence of his citation analysis. ‘In Italy, the SCI was like salvation to some scientists’ for the way it could highlight the ‘unfair allocation of credit and resources’ (Hargittai, 1999, p. 26).

‘If the SCI is used in tenure evaluations, hopefully it is done intelligently’, Garfield explains (Hargittai, 1999, p. 26). Some institutions have been known to measure the length of the indexed citations in inches for an individual scholar’s tenure case (Stein and Daniels, 2016): more inches (in citations), more impact. This runs counter to what Garfield has advocated throughout his career: that his ‘automatic and objective’ measures should be used in combination with peer evaluation (Garfield and Malin, 1968, p. 7). Still, academic institutions with a management mindset find the promise of ‘automatic and objective’ measures irresistible.

Building on the SCI and SSCI, Garfield also created the ‘journal impact factor’ (Garfield, 2005). The journal impact factor (JIF) figures prominently in the UK’s measurement system for academic performance, the Research Excellence Framework (REF). The way it is used in most academic institutions, and the way it works in British universities, the JIF is regarded as a valid and reliable measure of scholarly impact. The JIF rates journals as a whole – not individual articles. To get this number, Garfield constructed a formula that analyzes the most recent one or two years of citations to articles in the journal (Garfield, 2006). As with many quantitative measures, such as a random-dial survey, the JIF provides a snapshot of a moment in time – in this case, the most recent two years. The JIF includes all citations to articles in a journal, including citations to articles published by the journal itself. This means that editors who want to raise their impact factor can do things to manipulate it, including publishing review articles that already include a large number of citations from that journal, both of which translate into a higher rating for the journal. It is common practice among journal editors, once an article is accepted for publication, to ask authors to add a few citations from that particular publication, known as ‘coercive citations’. This practice is so commonplace, writes one professor of engineering, that scholars anticipate what is expected and ‘load their articles with citations from the journal to which they are submitting before they are even asked’ (Hoole, 2014).

While some regard the JIF as one of the highest standards of peer-review measures, critics contend that it is little more than a measure of popularity and takes a very short-term view of impact. Critics

contend that this means that the JIF provides a misleading indication of the true impact of journals, biased in favor of journals that have a quick impact, as is the pattern in the sciences, rather than a prolonged impact, as happens more often in the social sciences and humanities (Vanclay, 2008). One analysis finds that the journal impact factor is ‘bad scientific practice’ as a way to measure scholarly impact (Brembs et al, 2013). This is not simply a critique of Garfield’s tool, but rather their data suggest that ‘any journal rank (not only the currently-favored Impact Factor) would have this negative impact’. Instead, they suggest abandoning journals altogether, in favor of a library-based scholarly communication system, which would use information technology to vastly improve the filter, sort and discovery functions of the current journal system (Brembs et al, 2013).

Some are so frustrated with the tool that Garfield created that they are taking action to have the JIF eliminated from any and all evaluation measures. In 2012, a group of scholars, journal editors and publishers, scholarly societies, and research funders across a range of disciplines issued a declaration calling on the world scientific community to stop using the JIF in evaluating research for funding, hiring, promotion, or institutional effectiveness (Stein and Daniels, 2016). To date, more than 12,500 individuals and 585 organizations have signed on to the San Francisco Declaration on Research Assessment (DORA).¹

‘When we talk about intellectual impact, it is very subjective’, concedes Garfield, although the tools he invented are widely regarded as being objective measures of impact. Of course, the evaluation of scholarly impact is more complicated than simply assigning a number. The assessment of scholarly impact is a complex evaluative process (Lamont, 2009). The process assumes an elite circle of readers deemed knowledgeable enough to assess quality and excellence. The review process in decisions about tenure, promotion, and grant awards is a deliberative one, in which the terms ‘quality’ and ‘excellence’ are hotly contested in the process (Lamont, 2009; Lamont and Huutoniemi, 2011). How scholars evaluate each other’s research varies, based on setting (internal review for hiring, tenure, and promotion), institution (elite research-focused universities or teaching-focused community colleges), and national context (Lamont, 2012). Quantitative indices, like the ones Garfield developed, are meant to serve as proxy measures for research quality and scholarly impact.

Garfield's tools and the impact agenda

There is an animated and nuanced discussion about how to measure scholarly impact in the UK, in large part because there is a funding structure for higher education there that relies heavily on the REF,² introduced in 2014. The REF emerged from a series of changes to higher education and other sectors in Britain, beginning in the early 1980s in the Thatcher years. Then began what many have now come to refer to as 'audit culture' and 'new managerialism', both expressions of a more global process of neoliberal economic and political transformation (Shore and Wright, 2003, p. 58).

With the Education Reform Act 1988, academics lost the security of tenure, and the funding of universities was tied to performance measures. Under royal charters, universities set their own standards and were the sole arbiters of their own quality. Audit culture, introduced by the Education Reform Act, marked a significant break with the principle of academic autonomy (Shore and Wright, 2003, p. 70). It laid the groundwork for the REF, and set current standards for judging scholarly impact in the UK.

The shift to audit culture in British higher education required departments to submit 'bids' claiming their provision to be 'excellent,' 'satisfactory' or 'unsatisfactory' (Shore and Wright, 2003, p. 70). Any university department deemed unsatisfactory had to rectify the situation within 12 months, or else 'core funding' would be withdrawn' (Shore and Wright, 2003, p. 70). The changing policies for higher education in the UK, writes John Holmwood, British sociologist and former head of the British Sociological Association, subordinates higher education to the market in ways that systematically undermine the idea of a public university and education for all (Holmwood, 2011). The rise in audit culture, and with it the push to measure impact, is part of the larger trend toward seeing higher education as a private commodity rather than a public good (Holmwood, 2011).

The meaning of 'teaching quality' has also been transformed by the audit culture of the neoliberal university. To be audited, the learning experience must now be quantified and standardized so that it can be measured. The curriculum's merits are today measured in terms of finite, tangible, transferable and, above all, marketable skills (Shore and Wright, 2003, pp. 72–3). If students are merely customers, then it becomes easier to shift the burden of paying for higher education onto them, while placing the burden of demonstrating 'satisfactory' impact onto faculty and a new army of academic administrators.

In the US, there is a fairly wide discussion within academia about scholarly engagement, prompted by Boyer's work, but compared to the UK, there is relatively little discussion beyond a few circles in the US about how to measure scholarly impact in bibliometrics or scientometrics, the areas of study that Garfield helped to launch. Rather than talk about an 'impact agenda', as is common in the UK, an analogous discussion in the US might be called the 'success agenda,' and it is usually framed by terms like 'teaching effectiveness', student 'outcomes' or the 'success' of students. The 'success agenda' in the US is led most vociferously by those advocating for charter schools, which are pre-college learning institutions that are also being transformed into corporate-style revenue generators (Fabricant and Fine, 2012). Those championing the success agenda in higher education in the US are those who pose questions such as: 'how well do graduates do a decade after their degrees? What do graduates [actually learn]?' (DeMillo, 2015, p. 17). Such questions are supposedly meant to promote measures of impact that are 'student-centered' rather than 'faculty-centered' (DeMillo, 2015). The suggestion to do away with research in universities is not truly one that serves the interests of students, who are often involved in and learn from research, but rather one that is economically centered but disguised as being student-centered.

There are no reliable comparative data on the use of metrics and methods of evaluations internationally (Lamont, 2012). For the two cases with which we are most familiar, the US and the UK, the underlying logic for the way 'impact' is understood is strikingly similar. Both the impact agenda in academia in the UK and the success agenda in the US are driven by the logic of the neoliberal marketplace. In the UK, the impact agenda is part of a broader audit culture and, in the US, the success agenda is part of the relentless focus on 'student outcomes' and 'achievement gaps.' Such 'impact' and 'success' agendas share a fundamental view of research and education as only worthwhile when measured in economic terms. There are differences, too. In the US, the funding for higher education is steadily and sharply declining without regard to measures, whereas in the UK, funding for higher education overall is somewhat more assured, although the distribution of it is tied to performance measures and increasing burden is placed on students to underwrite the costs.

In response to the increasing pressures of audit culture and the rising tide of metrics in academia in the UK, some researchers are pushing back (Wilsdon et al, 2015). In June 2014, two British professors of international politics, Meera Sabaratnam and Paul Kirby, wrote about why metrics are inappropriate for assessing research quality (Sabaratnam

and Kirby, 2014). They argue that the metrics used in the REF, including Garfield's journal impact factor, are not accurately measuring research quality. They also point out that such a system 'systematically discriminate[s] against less established scholars', who are less likely to have high citation counts. Relying on citation indices, like the ones Garfield created, also disadvantages work by women and ethnic minorities, because their work is less likely to be cited (Sabaratnam and Kirby, 2014). They note that the putatively 'objective' measures like citation counts are 'highly vulnerable to gaming and manipulation' through practices like coercive citations (Sabaratnam and Kirby, 2014). The overall effects of using citations as a proxy for either 'impact' or 'quality' would be 'extremely deleterious to the standing and quality of UK academic research as a whole', they conclude. While Sabaratnam and Kirby report an overwhelmingly positive response to their activist blog post, there are other, more damaging unintended consequences from audit culture.

There is growing evidence that the emphasis on metrics is having a crushing effect on some academics. Scholars can feel an enormous individual pressure to produce publications and to win grants in order to satisfy the metrics of impact at their institution. In one tragic case, the emphasis on measuring scholarly impact cost the life of Stefan Grimm. A well-regarded and well-published scholar in toxicology at Imperial College London, Grimm committed suicide after being placed on 'performance review' following the news that his grant applications did not get funded (Parr, 2014). The reliance on a narrow range of metrics is prompting those scholars who want to have children to consider what this will mean for their research track records. In the context of an increasingly neoliberal university, where every activity is audited, some raise the provocative question: 'How many papers is a baby worth?' (Klocker and Drozdowski, 2012). To be sure, these are symptoms of larger global processes of neoliberal economic and political policies and not merely the result of Garfield's citation analysis tools. But his tools have had some profound, if unintended, consequences for measuring scholarly impact in the academy.

For over three decades, ISI, under Garfield's leadership, produced and marketed a wide range of information management tools. In 1988, Garfield sold ISI to a mid-sized publishing company for a reported \$24 million. Then, in 1992, ISI was sold again for \$210 million to multi-billion-dollar media conglomerate Thomson Reuters (Lane, 1992). Since he developed them in the middle of the 20th century, Garfield's citation analysis tools, originally intended to spot scientific trends and map the flow of ideas, have become widely institutionalized

to evaluate scholarly impact. These tools have then become privatized and monetized, most recently by Thomson Reuters. What this means is that now, in order to access these tools, academic libraries pay a hefty fee to Thomson Reuters for academic review committees to be able to use them.

The dominance of Garfield's tools in measuring scholarly impact disrupted a previous system of evaluation that relied solely on reputation and personal relationship. This resulted in institutions filled with 'administrators instead of world-class scientists', as Garfield explained (interview by Istvan Hargittai, October 1999). Yet, the system of citation analysis and indices that he developed is subject to manipulation through a variety of means, including coercive citation practices. When used on its own and not in combination with peer evaluation, citation analysis can reproduce existing hierarchies within academia by systematically disadvantaging everyone except white men from elite institutions, who are more likely to cite themselves and each other.

Garfield's tools are also effectively behind a paywall at Thomson Reuters; knowingly or not, academics 'pay' to use them as a benefit of their university library affiliations. Now, the dominant practices that have developed in tandem with Garfield's citation analysis tools that are being disrupted by digital media technologies.

As the application of Garfield's tools suggests, new technologies and open scholarship cannot disrupt traditional patterns of citation without a steady intention to shift the infrastructure of our scholarly systems. Veletsianos and Kimmons argue that 'merely developing digital literacies ... does not mean that scholars will necessarily become efficient or equal participants in online spaces' (Veletsianos and Kimmons, 2012). We must apply open scholarship and digital practices to circulate ideas from the scholarly margins as well as those from the disciplinary mainstream. We must aim to create rich and varied open resources and practices to support intersectional scholarly work that unsettles methods and assumptions. We must conscientiously critique our scholarly tools and assessment patterns. Without new ways of assigning value to scholarship, we will perpetuate dominant systems of thought at the expense of innovation and creativity. The conditions that constrain attention to some ideas and support attention for others cannot be addressed with technology alone; they require social and political adjustments.

The flow of scholarly information is becoming more open. The fact that academic work, even that behind paywalls, is searchable by Google Scholar at relatively low cost is an important shift and one that promises to reconfigure evaluative practices in the academy (Lamont,

2012). We are just at the beginning of understanding how digital media technologies will transform the way we measure scholarly impact in the academy. When asked about how the Internet was changing scholarly impact, Garfield replied with another question of his own: ‘The Internet is having an impact but how would you measure it?’ (Hargittai, 1999, p. 26).

Digital media and the rise of altmetrics

Jason Priem and an international group of collaborators have some ideas about how to measure scholarly impact in the digital era. In 2010, Priem, along with Dario Taraborelli, Paul Groth, and Cameron Neylon, drafted a ‘manifesto’ calling for ‘more tools and research based on altmetrics’ (Priem et al, 2010). Altmetrics, as they explain, is an alternative form of metric – hence, ‘altmetric’ – that reflects the changing flows of scholarly information in the digital era better than Garfield’s citation analysis. They explain the need for altmetrics this way:

No one can read everything. We rely on filters to make sense of the scholarly literature, but the narrow, traditional filters are being swamped. However, the growth of new, online scholarly tools allows us to make new filters; these altmetrics reflect the broad, rapid impact of scholarship in this burgeoning ecosystem. We call for more tools and research based on altmetrics. (Priem et al, 2010)

Rather than the journal-based metric of Garfield’s JIF, altmetrics enable article-level measurement, which better reflects the way scholars search and find information. Typically, we find, read, and use an article based on discovery through specific search terms, and altmetrics can reflect that. Unlike the citation metrics that Garfield developed, altmetrics will track the spread of ideas outside the academy, including circulation in sources that are not peer-reviewed. Priem and colleagues argue that altmetrics are less susceptible to manipulation than the JIF. In effect, they suggest using the statistical power of big data to algorithmically detect work that is being cited across a wide array of platforms, not only in academic journals. Some academics are excited about altmetrics, in the hope that this new set of metrics will provide an innovative way to measure scholarly impact (Matthews, 2015).

‘If you want people to find and read your research, build up a digital presence in your discipline, and use it to promote your work when you have something interesting to share. It’s pretty darn obvious, really’, suggests Melissa Terras. Terras trained in art history, her research focuses on using computational techniques to enable research in the arts and humanities that would otherwise be impossible. She became curious about the way social media is affecting the dissemination of her work, and decided to conduct an experiment.

Terras had posted her papers in her institutional repository, but she could see that most had only one or two downloads. She decided to find out what would happen if she blogged and tweeted about them. Terras discovered that her scholarly papers that were mentioned on social media had at least more than 11 times the number of downloads than similar papers that were not. Upon blogging and tweeting, within 24 hours, there were, on average, 70 downloads. One paper of hers in particular was downloaded over a thousand times in the year following her social media experiment. Her paper became the sixteenth most downloaded paper in the entire institutional repository in the final quarter of 2011. This paper was also the most downloaded paper in 2011 in *LLC Journal*, where it was published, with 376 full-text downloads (Terras, 2012).

Terras’ experiment, with her 376 full-text downloads, is an example of the different approach to measuring scholarly impact that is implicit in altmetrics. She demonstrated the power of social media to boost readership of her work and she has data to support this. Her experiment also highlights the article as the unit of analysis, typical of altmetrics, rather than the journal, as with the JIF. Terras’ experiment was conducted in the spirit of Boyer’s notion of the scholarship of engagement, and a clear example of the scholarship of sharing knowledge. She wanted people to find her work and read it – and they did. But what Terras’ experiment and the rise of altmetrics do not yet do is to reach back into the academy.

At the time of writing, we do not know of any colleges or universities that officially use altmetrics in the tenure and promotion process. However, traditional journals incorporate evidence from altmetrics of wider readership into their journal’s online interface. Grant funders are very interested in alternative measures of influence and reach that incorporate social media data.

In many ways, the rise of altmetrics speaks to the kind of change that Kathleen Fitzpatrick evokes concerning peer review in the digital era: that it is transforming from a process focused on gatekeeping to one concerned with filtering the wealth of scholarly material made

available via the Internet (Fitzpatrick, 2010). In other words, instead of producing knowledge in a context in which knowledge is rare and hard to access, now scholars are creating knowledge in a context of abundance and information overload (Shenk, 1997; Stewart, 2015). Altmetrics is a way to measure impact that takes into account these new filtering systems, like search engines and article downloads. Whether or not altmetrics will truly disrupt Garfield's citation analysis tools remains to be seen, but Garfield's career is still relevant here.

The story of Eugene Garfield's unintended influence on how we measure scholarly impact may turn out to be something of a cautionary tale for the digital era and the rise of altmetrics. 'After five years, we still don't have much of an idea of what we're measuring', Juan Pablo Alperin told a room full of altmetrics enthusiasts at a conference in Amsterdam (Matthews, 2015). Alperin, a professor at Simon Fraser University in Canada, is an expert in online scholarly communications and how we measure it. Alperin voiced concerns that altmetrics might lead to the creation of a new, all-encompassing metric that merely replaced citation-based measurement. 'Aren't we going down the same route [with altmetrics]?', he asked (Matthews, 2015). If altmetrics becomes enclosed and monetized, then it will be going down the same route as Garfield's citation analysis tools. To escape from this same route, we have to rethink engagement.

From impact to engagement

Returning to Boyer's concept of scholarly engagement and Stewart's analysis of this for the digital era might suggest how we might think differently about scholarly impact. Stewart contends that digitally networked scholarship makes a difference in how scholarly communities, whether institutions or entire disciplines, understand and honor the values of scholarly inquiry that presume a context of scarcity of knowledge, when scholarship is now being done in a context of information abundance (Stewart, 2015). If abundance and openness are the context of contemporary scholarship, and Boyer's typology is the guiding principle for how we think about impact, then engagement might look very different.

Open syllabus metadata

The citation analysis tools SCI, SSCI and JIF only function within the universe of journal articles and cannot be used to measure the impact of books or other kinds of scholarly work. If one were to imagine a broader view of scholarly engagement that included books and found a way to measure how widely read (or at least assigned) they were across college campuses, then you might have something like the Open Syllabus Project.³

The Open Syllabus Project finds course syllabi available on the open web and looks for which books are being assigned in US college classes. To date, the project has collected over one million syllabi, and has extracted citations and other metadata from them. The Open Syllabus team collects metadata – there is no individual or personally identifying information in their database. It is open for everyone to explore. ‘Such data has many uses. For academics, for example, it offers a window onto something they generally know very little about: how widely their work is read’, says one of the developers behind the project (Karaganis and McClure, 2016).

The Open Syllabus Project allows for a new kind of publication metric based on the frequency with which books are taught, which the developers are calling a ‘teaching score’: ‘The score is derived from the ranking order of the text, not the raw number of citations, such that a book or article that is used in four or five classes gets a score of one, while Plato’s *The Republic*, which is assigned 3,500 times, gets a score of 100’ (Karaganis and McClure, 2016). The results from their initial analysis are promising in terms of the wider range of voices represented in the classroom. In US courses covering fiction from the last 50 years, the most frequently taught book is Toni Morrison’s *Beloved*, followed by William Gibson’s *Neuromancer*, Art Spiegelman’s *Maus*, Toni Morrison’s *The bluest eye*, Sandra Cisneros’ *The house on Mango Street*, Anne Moody’s *Coming of age in Mississippi*, Leslie Marmon Silko’s *Ceremony* and Alice Walker’s *The color purple* (Karaganis and McClure, 2016).

The developers behind the Open Syllabus Project are clear in tying their project to the idea of scholarly engagement and a different way of measuring impact:

If you like the idea of a more publicly engaged academy, you need to look elsewhere for incentives. That’s where we think our ‘teaching score’ metric could be useful. Teaching captures a very different set of judgments about what is

important than publication does. In particular, it accords more value to qualities that are useful in the classroom, like accessibility and clarity. A widely taught but infrequently cited article is an important achievement, but an invisible one to current impact metrics. (Karaganis and McClure, 2016)

If having students in college classes engage with your work is one priority for rethinking scholarly engagement, so, too, is reconsidering the relationship between universities and the communities that surround and support them.

University–community engagement

Kristine Miller is interested in design and neighborhoods, but she found the usual approach lacking. ‘A lot of times in design programs, people tend to improve places sort of expecting “new and better” people to move in, while our question was, how can our professions make neighborhoods better for the people who live there currently?’ (Miller, quoted in Hinterberg, 2015).

Miller is a professor at the University of Minnesota, where she experiments with ways to open scholarly impact to other kinds of evaluation in line with Boyer’s scholarship of engagement. The University of Minnesota is developing a set of Public Engagement Metrics for their faculty. These include measures such as: evidence of application of research findings by communities and evidence of the contribution of public engagement to student learning. As part of this initiative, the University of Minnesota gives awards to faculty who are doing exemplary work in community engagement, including Kristine Miller, who received an award in 2015.

In 2005, Miller began a collaboration with Juxtaposition Arts as part of a partnership called Remix, with the goal to make the design of cities more equitable for the people who live in them: ‘I want all of my students to find that sweet spot between what they’re really good at and the change that they want to see in the world. And I think that engaged scholarship is a really way good way of doing that’ (Miller, quoted in Hinterberg, 2015). When reflecting on the success of the project, Miller talks about relationships:

Our success can really be measured in the people who are coming out of our program and the relationships that we’ve

developed. It's exciting now, 10 years into it, to see the next generation, to see students who have come through our program taking on roles in the Twin Cities and elsewhere drawing attention to equity issues. (Hinterberg, 2015)

The kind of work Miller is doing is a collaborative project, one that involves a large number of people – both students and people in the community – over a long period of time. In traditional measures of scholarly impact, Miller's community engagement work would be discounted in favor of the number of citations she has and the impact factor of the journals in which she has published. Shifting to a paradigm of engagement, and finding ways to reward it, makes this important kind of work legible within the academy.

Transactional and transformational metrics

Transactional metrics include both traditional academic measures of impact (such as citations), and alternative measures (like altmetrics such as downloads or mentions on social media). Quantitative transactional measures of impact can also include lasting social change, such as changes to public policy. One of the useful aspects about this conceptualization is that it illustrates the incremental change that altmetrics represent. In other words, altmetrics are just another way of counting things – downloads and social media mentions, rather than citations – but it's still just counting things. Counting and quantification can tell us some things, but it doesn't provide a whole picture of scholarly engagement.

On the 'transformational' side are those things that it is difficult, perhaps even impossible, to measure, but that are so crucial to doing work that has a lasting impact. These include identifying allies, building relationships, establishing collaborations, and co-creating projects. Ultimately, transformational work is about changing lives, changing the broader cultural narrative, and changing society in ways that make it more just and democratic for all. These kinds of transformations demand a different kind of metric, one that relies primarily on storytelling. How might this work in academia? Well, to some extent, it already does.

To take the example of teaching, you may have got the advice 'save everything' for your tenure file. This advice often goes something like, 'every time a student sends you a thank you card, or writes you an email, or says, "this class made all the difference for me", save that for your tenure file'. That's part of how we 'finish the symphony', to

borrow Doug Howard's metaphor; we get notes from students and we compile those into a narrative about our teaching (Howard, 2014). It's impartial, to be sure, but it's something. The comments that students add to teaching evaluations are another place where we see impact in narrative form,⁴ although these are so skewed by the context of actually sitting in the class that it misses the longer-term impact of how that course may have changed someone.

For the diminishing few with tenure, those letters of recommendation we write in support of junior scholars are another example of the use of narrative in evaluation. Whether we are writing for someone to get hired, promoted, or granted tenure, what we are doing when we craft those letters is creating a narrative about the candidate's impact on their corner of the academic world so far. Of course, we augment that with quantitative data, 'this many articles over this span of time', and 'these numbers in teaching evaluations'. The fact is, we already combine transactional and transformational metrics in academia in the way that we do peer evaluations. What we need to consider in academia is expanding how we think about 'impact' and realize the way that we already use both quantitative and qualitative measures to evaluate and assess the impact of our work.

Our (mostly failed) experiment with metrics

Given the context of the neoliberal university, raising the subject of metrics can set off alarm bells for faculty, who are right to be concerned about measures of output being used to demolish scholarship and subvert tenure (Holmwood, 2011, 2014). We ran into some of this resistance when we tried to incorporate a survey about metrics with our project.

Our idea was to do a simple pre- and post-test assessment of the increase (we hoped) in social media use and mainstream media mentions of faculty research. We would do this by sending a brief survey to the faculty, ask about their engagement with social media, and then track both social media use and mainstream media mentions. As an academic institution, ours is quite unique in structure. There is a small cohort of faculty (around 150) appointed exclusively at the Graduate Center, which is the PhD-granting arm of CUNY, and nearly all the faculty in the disciplines have tenure. Many are 'distinguished' faculty (the rank above full professor), and most are very prominent scholars in senior stages of their careers. There is a much larger group of doctoral faculty (around 1,800), who have a primary appointment at one of the 24

campuses of CUNY and also have joint appointments at the Graduate Center. This consortial model of a university means many things, but one artifact of our organizational structure is that there is no way to send an all-faculty email across the consortium.

Our plan was to begin by surveying just the smaller group of central faculty, then extend the survey to the affiliated doctoral faculty. Informal discussions with faculty about the idea of the survey raised concerns because our project had the support of the Provost's office. A number of faculty members balked at the mere suggestion of completing such a survey; another faculty member suggested that this was an indication of a larger trend of 'metrification' in higher education and wanted nothing to do with it. Still another faculty member predicted a low response rate on anything coming from the Provost's office. Internally, a debate ensued about who might send out the survey to minimize the fact that it did have the endorsement of the Provost. Eventually, we abandoned the idea altogether and chalked it up to one of the many lessons learned about the limits of top-down change in an institution with an empowered faculty.

At the same time we met with this resistance, we were busy training scholars in the MediaCamp workshops in a variety of hybrid, digital skills (see Chapter Five). The people who came to our workshops evidenced a larger trend within academia. Increasingly, scholars are using digital media technologies to do their work, collaborate and find others who share their research interests (Lupton, 2014; Carrigan, 2016a). Scholars who are immersed in digital media technologies expect this work to count when it comes to decisions about hiring and tenure (Starkman, 2013; Matthew, 2016). The reality that digitally fluent scholars confront in most institutions is one in which legacy measures of scholarly impact are deeply entrenched. At this time, there are no good ways to measure scholarly engagement in ways that are equitable across disciplines and legible across all institutions of higher education.

Forward thinking: the humane ends of scholarship

We began this chapter with a discussion of Boyer's scholarship of engagement. For Boyer, universities and colleges were among 'the greatest sources of hope for intellectual and civic progress' (Boyer, 1996). We tend to agree. There is so much possibility in the work of scholars to address social inequality in all its dimensions. In discussing the scholarship of engagement, Boyer said: 'The issue, then, ... is not

whether scholarship will be applied but whether the work of scholars will be directed toward humane ends' (Boyer, 1996). In our view, it is this idea of the 'humane ends' of scholarship that may provide a way forward in thinking about how to measure scholarly impact through a metric not based on citation analysis but on social justice.

'In the century ahead, higher education in this country has an urgent obligation to become more vigorously engaged in the issues of our day' (Boyer, 1996, p. 28). On campuses like the University of Missouri this past year, the 'most pressing issues of the day' has been #BlackLivesMatter. The movement began as a hashtag to raise awareness about the death of Trayvon Martin and the extrajudicial killing of black people across the US. When a recent high school graduate, Michael Brown, was shot and killed by a white police officer in Ferguson, Missouri, and his body left on a hot, August pavement for hours, protestors took to the streets and mobilized online using a variety of hashtags including #BlackLivesMatter, #Ferguson and #MikeBrown. At the nearby University of Missouri, just an hour-and-a-half away by car from Ferguson, students began 'MU for Mike Brown', a Black Lives Matter-affiliated group formed in solidarity with the uprisings over the shooting to death of an unarmed teenager.

Solidarity with the #BlackLivesMatter protest group gave rise to a second group, Concerned Student 1-9-5-0, a reference to the year 1950, when black students were first admitted to the University of Missouri (Izadi, 2015). These students were spurred to action by a pattern of racial harassment they experienced on their campus, such as a swastika drawn in excrement on a campus building, shouts of the 'n-word' yelled from passing cars at African American students, and a Legion of Black Collegians theater rehearsal disrupted by more racial slurs being yelled, along with a constant drumbeat of other incidents. In October, during a homecoming parade, students blocked University President Tim Wolfe's car, insisting that he respond to the series of racist incidents on campus. Wolfe expressed concern, but did not leave his car. A graduate student, Jonathan Butler, released a list of demands and began a hunger strike. At the top of the list of demands was that Wolfe resign, and that any replacement be chosen through a collaborative process. Also among the demands were: 'We demand that by the academic year 2017-2018, the University of Missouri increases the percentage of black faculty and staff campus-wide to 10%' and develop a strategic plan to 'improve retention rates' among marginalized students by changing the campus environment (Concerned Student 1-9-5-0, 2015). The predominantly African American football team, along with white teammates and coach, also joined the calls for the president to

resign and threatened not to play until he did so (Green, 2015). If the football players went on strike, the University of Missouri would stand to lose millions in football-generated revenue. On November 9, Tim Wolfe resigned (Eligon and Pérez-Peña, 2015).

Part of what the students at the University of Missouri did was to change the conversation about metrics to ones that mattered to them. Instead of the metrics about productivity or student success, they started a different conversation about social justice metrics. The relevant metrics had to do with the white dominance of higher education, and their experience of campus life in particular, such as the number of times they got harassed on campus based on race or ethnicity (frequently), and the number of faculty who are people of color (fewer than 10%). Within academia, the issue of the white dominance of higher education most frequently gets expressed as a 'lack of racial diversity among faculty', but this does not go nearly far enough to describe the current situation. Colleges and universities who want to fulfill Boyer's vision and be engaged in the 'most pressing issues of the day' are going to face an uphill struggle, given the pattern of systemic racism in US higher education.

Boyer's vision of colleges and universities as 'the most promising institutions for progress' is one that was premised on his narrative about land grants (issued by President Lincoln 'during the dark days of the Civil War' to help 'workers and farmers') and the GI Bill ('a wonderful experiment' of 'rising expectations' (Boyer, 1990)). In fact, higher education in the US has never been a level playing field (Herbold, 1994). Although some 8 million veterans were able to take advantage of the GI Bill to access a college education, most African American veterans were not able to take advantage of this federal program to which they were fully entitled due to rampant discrimination on college and university campuses (Herbold, 1994). This was not a new development in the post-World War II era. As historian Craig Wilder chronicles in his book, *Ebony and Ivy* (2013), the leading universities in the US were built using the labor of enslaved people and were dependent on human bondage for their operation through the first half of the 19th century (Wilder, 2013). Understandably, this affected the production of knowledge at these same institutions as they became the sites for the development and sustenance of racist ideologies. The question before us is whether or not the university can become a place 'vigorously engaged in the issues of the day' when those issues are at the very heart of the inequality within the university itself.

Empowered, digitally networked faculty and students pose a serious challenge to the smooth management of the neoliberal university. Tim

Wolfe, in some ways, typifies the new management of the neoliberal university. He came to the University of Missouri from private industry, starting his career at IBM and eventually becoming president of Novell, a software company (Wilson, 2015). Part of what students (and some faculty) did at the University of Missouri was to use digital media technology to disrupt the business-as-usual operation of the university, much like ACT UP protestors did in the 1980s (see Chapter Four). They used the ‘internet empires’ that DeMillo warned make it difficult to govern a university, and in this, he is correct. Another component to their challenge to the neoliberal university was to effectively change the most relevant metrics at hand. This is part of what the students at the University of Missouri and other campuses did in the fall of 2015 – they changed the metrics being discussed. If we reimagine a world in which higher education is engaged in trying to end inequality rather than continuing as an engine of it, then we might imagine different kinds of metrics, ones that assess how a particular institution is addressing the social inequality around it.

We return to where we began this chapter, with Boyer’s call to a scholarship of engagement. For Boyer, the call to be engaged scholars is not an end in itself, but one that propels us toward the ‘humane ends of scholarship’ (Boyer, 1996). It is this idea that may provide a way forward in thinking about how to measure scholarly impact through a metric not based on citation analysis but on social justice. However, we recognize that such a reconceptualization of metrics is wholly at odds with the logics of the neoliberal university and the politics of austerity. But it is possible to confront these, as the students at the University of Missouri have taught us. We would argue that it is imperative that we follow their lead for the sake of higher education.

In many ways, the student protestors and their list of metric-embedded demands are making colleges and universities more ‘vigorously engaged in the issues of our day’ – to use Boyer’s phrase – and more politically relevant than they have been since the 1960s. Jennifer Wilson, a postdoctoral fellow at the University of Pennsylvania, put it this way: ‘At a time when the value of a college education, particularly a humanities education, is under attack, the Black Lives Matter campus protestors of 2015 have shown, through their own display of historically informed and intersectional politics, that universities matter’ (Wilson, 2015). To move toward the humane ends of scholarship and be vigorously engaged in the issues of the day, we urgently need to change the conversation about metrics to one that enables us to think about the university in relation to social inequality.

Notes

¹ See www.ascb.org/dora

² See www.ref.ac.uk

³ See <http://opensyllabusproject.org/>

⁴ Of course, the use of student evaluations as reliable metrics is contentious. The comments from disgruntled students can have a negative impact on faculty careers. Research has consistently shown that the race and gender of the professor have a significant impact on students' assessment of faculty. See, for example, Anderson and Smith (2005).