Title
New Metrics for Measuring Academic Research Outside the Ivory Tower

ISBN
978-1-78973-270-2

Author
Levine, Kendra K.

Publication Date
2019-08-01

Berkeley Institute of Transportation Studies

New Metrics for Measuring Academic Research Outside the Ivory Tower

The New Metrics: Practical Assessment of Research Impact, Emerald Publishing Limited, pp. 39-58.

DOI: 10.1108/978-1-78973-269-620191003

Kendra K. Levine, ITS Library, Institute of Transportation Studies, University of California, Berkeley

August, 2019

Berkeley UNIVERSITY OF CALIFORNIA

New Metrics for Measuring Academic Research Outside the Ivory Tower

The need to demonstrate the value of research programs is critical for funding sustainability, particularly in the 21st century where research funding is rarely guaranteed beyond the life of a single project. After the great economic crash of 2008 many research programs have also diversified their funding portfolios, which means cultivating new relationships with funders and potentially different objectives of research projects. Measuring the impact of academic research outside academia is critical to providing a holistic view of a research program, but the methods to do so are still being figured out and are largely context specific. Within the field of transportation, no metrics for measuring the impact of research have really been established. Readily available and easily measured metrics, like citations or other scholarly impact factors are not entirely appropriate to assess these kinds of research, since it hews to a narrow academic setting. This chapter proposes a methodology to track and measure the impact and research of beyond academic settings across a number of academic disciplines related to transportation conducted by the University of California Institute of Transportation Studies (UC-ITS), a multi-campus research organization at UC Berkeley, UC Davis, UC Irvine, and UCLA. The methodology is designed to be adaptable for use by other research groups that straddle academia and the public sector, regardless of the disciplines involved.

Research Impacts

Established Scholarly Metrics

Traditional scholarly metrics, for all their ills, have been established to describe the impact of academic research within academic venues. These methods can be considered acceptable for academic evaluation, like promotion and tenure cases, academic research program reviews, or grant applications. Scholarly metrics like journal impact factors are insufficient for demonstrating the impact of research beyond academia, especially since the mechanism to accept and publish research papers in academic journals (peer review) can be seen "as much of a lottery as a rational process" (Segeln, 1997). Throughout the 1990s several researchers explored how journal and article citations could be used to evaluate research, though the tools available at the time were limited due to the maturity and sophistication of citation databases at that point (Taubes, 1993). In the intervening years a lot of progress has been made in developing systems and platforms to track citation data, but there is still a need for an open ecosystem for this data for more robust analysis outside vendor platforms or questionable black boxes like Google Scholar (Shotton, 2013). The Journal Impact Factors (JIF) is still frequently cited and relied upon despite established limitations in measuring anything about research beyond a raw number of citations (Segeln, 1997). There is merit to discussions about the possibilities of abandoning JIFs for evaluation purposes and developing a new, more deliberate and thoughtful approach for scholarly metrics, but JIFs are so entrenched that there is little chance of that happening at scale (Archambault & Larivière, 2009). The PLoS Medicine Editors published an editorial in 2006 about the potential gamification of impact factors, concluding that as scientific publishing is democratized and authors and readers have more options to engage outside the established (and tracked) publishing platforms, JIF will be less relevant. "If authors are going to quote the impact

factor of a journal, they should understand what it can and cannot measure. The opening up of the literature means that better ways of assessing papers and journals are coming—and we should embrace them" (PLoS, 2016). Any continued reliance on these metrics needs to account for their clear limitations and known problems. JIFs need to be contextualized as a reminder of what they actually measure, what that can represent, but more importantly what is cannot.

Using the same methodology for evaluating research impacts across disciplines is problematic because coverage in Journal Citation Reports (JCR) in Clarivate, formerly known as ISI Web of Science, and other citation tracking databases is not uniform across fields. In mathematics, for example, many of the leading journals were not indexed by Clarivate (Korevaar, 1996). Normalizing citation data for bibliometric research is one way to address subject-based limitations for comparisons, though that is not frequently done when research groups or departments are evaluating themselves or their peers (Bornmann, 2013). Proposed new metrics based on these normalization techniques, such as Source Normalized Impact Per Paper (SNIP) or Field-Weighted Citation Impact (FWCI), could be widely adopted to provide a more fair evaluation tool across science (Kochetkov, 2018). However any new measures should only be used if they actually add value to the evaluation process and are not simply another number to be gamed (Watermeyer & Hedgecoe, 2016). Despite (or because of) the proliferation of academic journals in transportation, the impact metrics for transportation literature range from barely measurable to decently respectable through are still relatively small compared other fields like medicine (Banister, 2014). Comparing JIFs across disciplines lacks nuance and assumes that all fields of research should be held to the same arbitrary standard, which is in essence a fool's game.

For many fields the demonstrable impact of research as applied outside university settings is far greater than could be measured by JIFs, h-indexes, or other scholarly metrics. For fields of research that have direct and tangible benefits outside academia, as evident through policies, new operations, improved systems which have impacts on society, the environment, and economy, their impacts and reach extend far beyond academia. Comparing academic researchers and programs whose research is close to application and deployment solely using traditional scholarly metrics, such as bibliometric indicators, ignores these other metrics to evaluate the impact of research. One reason for the difficulty is frequently these research outputs are published as some form of grey literature, such as technical reports, white papers or policy briefs. As a result, their impact is more difficult to track and measure since they are usually excluded from most scholarly metrics systems, and many of their citations might be in other forms of grey literature, standards, or other kinds of working documents for different organizations. Google Scholar has proved to be useful in measuring the impact of these kinds of research outputs, by casting a wider net of tracked citations than Scopus or Clarivate (Haddaway, Collins, Coughlin & Kirk, 2015). Google Scholar citations have many limitations though, since their accuracy and consistency is far from guaranteed (Falagas, Pitsouni, Malietzis, & Pappas, 2007). Altmetrics are another tool to measure the potential reach of research outside academic settings, though there are still many issues to work out, such as differentiating between buzz of people talking about new findings and actual impact of how those findings are used in the field (Bornmann, 2014).

There is also the fundamental issue of defining "impact" in an academic research context (Terämä, Smallman, Lock, Johnson & Austwick, 2016). Different research disciplines have their own measures of impact, making any comparisons across them potentially contentious. Academics also have their own ideas of impact that likely differ from research funders and other

stakeholders. Faculty and graduate students on the academic job market need to publish their research in highly ranked academic journals that are valued by promotion and tenure committees. The journals are not necessarily read or by or even accessible to practitioners in the field who are involved in setting the research agenda, developing and advising on policy, and deploying new processes or technology into the field. Different publication venues and formats will reach and appeal to different audiences, and therefore must be accounted and tracked to fully assess the impact of a research project or program. As noted before, Clarivate and other scholarly metrics tools do not work for grey literature because they are not part of those ecosystems. By its very nature of grey literature is accessible (and diffuse) because it is widely available to the public on the web. This is important because practitioners by their nature by in large work for public agencies or private companies, not universities or colleges with access to scholarly publications through libraries. Thus accounting for citations in grey literature is necessary when looking at research impact beyond academia. For some fields, such as agriculture and aerospace, forms of grey literature are largely respected and expected to be used and cited in research (Auger, 1998). For public policy development and analysis, grey literature is a vital resource used in conjunction with scholarly sources (Lawrence, Houghton, Thomas & Weldon, 2014). Currently, tracking citations in that fashion is a very labor-intensive task in the vein of finding needles in haystacks. This is why people rely on Google Scholar even though it is a deeply flawed tool with questionable reliability.

Societal Impacts

In 1945, Vannevar Bush as head of the newly created Office of Scientific Research and Development (OSRD), issued the foundational report *Science: The Endless Frontier*, which outlined why public funding of scientific research was beneficial and valuable for society.

"The Government should accept new responsibilities for promoting the flow of new scientific knowledge and the development of scientific talent in our youth. These responsibilities are the proper concern of the Government, for they vitally affect our health, our jobs, and our national security. It is in keeping also with basic United States policy that the Government should foster the opening of new frontiers and this is the modern way to do it. For many years the Government has wisely supported research in the agricultural colleges and the benefits have been great. The time has come when such support should be extended to other fields" (United States., & Bush, 1945).

In addition to making the argument of the economic and societal benefits in sustained investment for basic research after the war effort, Bush called for the creation of a dedicated agency to manage and administer research. Bush's recommendation directly led to the creation of the National Science Foundation (NSF). Investment in scientific research in the immediate post-war period was an accepted good for national security and economic competition, but as resources for funding are no longer able to keep pace with research and innovation. Funding scarcity naturally led to increased scrutiny on the value of research and justification for continued funding. In 2001 Satler and Martin categorized the economic benefits of publicly funded research as:

- Increasing the stock of useful knowledge
- Training skilled graduates
- Creating new scientific instrumentation and methodologies
- Forming networks and stimulating social interaction
- Increasing the capacity for scientific and technological problem-solving
- Creating new firms (Salter & Martin, 2001)

It is notable that the focus on solving problems or improving existing systems is not one of their explicit benefits. Workforce development and fostering social cohesion through public knowledge do have economic benefits, but also clear societal benefits. Translating academic research to formats and venues that can be used by other parts of society is critical to ensure its relevance and thus give it a chance of having some impact outside the ivory tower. Nightingale and Scott suggest, "Funders need to recognize the distinction between relevance and academic impact" (Nightingale & Scott, 2007). They go on to argue that research that performs well according to traditional scholarly metrics might be good for an academic field but not society. "Impact may be easy to measure and audit, but relevance is not. The way that the thing being measured by impact metrics is changed by being measured (game-playing and so on) calls into question the entire foundation of the assumed association between research quality and the disciplinary judgements of value that inform performance metrics" (Nightingale & Scott, 2007). This concern about gaming the metrics hints at another problem in the relationship be academic research and society -- that it encourages researchers and funders to focus on areas that can be easily measured, often at the exclusion of less quantifiable research areas. Ernø-Kjølhede and Hansson warn, "the obvious danger is that researchers and universities intensify their efforts to participate in activities that can be directly documented rather than activities that are harder to document but in reality may be more useful to society" (Ernø-Kjølhede & Hansson, 2011). Accountability and assessment are important parts of any research program and should be part of the standard practice, but it cannot be discounted (and should likely be addressed elsewhere) that the desire for easily quantifiable metrics is shaping the way research is approached and likely detrimental to certain lines of inquiry. Developments in this area reflect the contemporary neoliberal research environment at public universities (and the rest of the public sector):

commodification and regulation (Burawoy, 2011). As a result of these pressures, the need and desire to show how research impacts and influences society is important, but there also needs to be an understanding of the roles and responsibilities of both sides. In 2016 Watermeyer distilled this important question about the focus on impact and regulation, and the actions of academics with the observation that academics and knowledge workers are "increasingly beholden to the strategic and technocratic frameworks" of research bodies and government funding agencies (Watermeyer, 2016). Societal impact of research is important and good, and should be considered when making decisions about public funding of research activities. As models of assessment are developed though, the focus on measurement "should not come at the expense of basic, blue-sky research, given that it is and will remain near-impossible to predict the impact of certain research projects years or decades down the line" (Bornmann, 2012).

One effort to assess the impact of research projects, and by extension research programs, is the Research Excellence Framework (REF). Developed in the United Kingdom by prominent higher education funding bodies in 2014 for a "shared policy aim for research assessment is to secure the continuation of a world-class, dynamic and responsive research base across the full academic spectrum within UK higher education" (Higher Education Funding Council for England, 2018). The REF provides accountability for public funding of research by producing evidence of the benefits of that investment, inform decision makers in the allocation of research funding. The REF is a process of expert review, systematically conducted by panels of academics and research users to assess the impact of research outputs outside academic settings (Higher Education Funding Council for England, 2018). During the REF 2014, 1,911 impact templates were assessed and graded by these review panels, with a total of 6,075 impact case studies (Manville et al, 2015). Developing and assessing the case studies and assembling the

panels to do so was very resource intensive and costly for all involved. The REF 2014 was seen as a worthwhile and successful exercise, somewhat unexpected for an initial effort, and the buy-in from academics and research users was pivotal in the REF's success (Manville et al, 2015). Critiques of the REF question if this culture of assessment is in conflict with the traditional missions of universities, and that rewarding and encouraging entrepreneurial academics will likely have long term effects not yet apparent, and the "overall impact of 'impact' on the academic community is thus too soon to calculate" (Watermeyer, 2016). If the REF is to really be successful and a culture of impact assessment is fully adopted by the UK academic research community, then there will also need to be dedicated resources (in funding and time) to make sure the initiative isn't "tokenistic" (Parker & Teijlingen, 2012).

The United States had a similar project that was launched in 2010; STAR METRICS® (Science and Technology for America's Reinvestment Measuring the EffecTs of Research on Innovation, Competitiveness and Science) is a project from the National Institutes of Health (NIH) and the NSF, under the auspices of the Office of Science and Technology (OSTP) (National Institutes of Health, 2019). The project reflected the goals of the American Recovery and Reinvestment Act (ARRA), the comprehensive federal stimulus bill focused on job creation, digital and physical infrastructure improvements, and increased government transparency and accountability. STAR METRICS® would look at the employment impacts of federally funded research, and then look at the broader impacts of federal scientific research along the following themes: Economic growth, workforce outcomes, scientific knowledge, and social outcomes (National Science Foundation, 2019). Collecting data to assess the programmatic impacts was halted in 2016 though, and the project has shifted focus with the development of Federal RePORTER, a database which will normalize data about federal scientific research grants

(National Institutes of Health (2), 2019). There is still a demonstrated need for a national framework for assessing and communicating the impacts of research programs.

Transportation

In transportation, frequently the impacts can be distilled to time, lives, and money. Time is often used to demonstrate different measures of mobility, such as how many vehicles travel on a stretch of road during a period of time, the travel times for system users, or how many people ride a transit system. The 2012 transportation authorization bill Moving Ahead for Progress in the 21st Century Act (MAP-21) established a performance- and outcome-based program to ensure that transportation projects were accountable and furthered national transportation goals. The seven performance areas targeted in MAP-21 follow the "lives, money, and time" model but are more detailed across different aspects of the transportation system. They are:

- Safety
- Infrastructure condition
- Congestion reduction
- System reliability
- Freight movement and economic vitality
- Environmental sustainability
- Reduced project delivery delays (Federal Highways Administration, 2013)

These performance measures were strengthened in the following transportation authorization bill, Fixing America's Surface Transportation (FAST) Act, which shortened the deadlines for State DOTs and MPOs to meet their performance goals, and if a state fails to meet their freight performance goals, they must include actions to correct that in their reports (Federal Highways Administration, 2017). This culture of performance measurement has extended to research, as

research programs (particularly those funded by federal and state agencies), and these areas of accountability are fundamental to those research programs.

The Environmental Protection Agency (EPA) uses 12 performance measures for evaluating sustainable transportation initiatives. They are similar to those in laid out in MAP-21 but also go a bit further to look at how transportation systems function and are planned. They are:

- Transit accessibility
- Bicycle and pedestrian mode share
- VMT per capita
- Carbon intensity
- Mixed land uses
- Transportation affordability
- Benefits by income group
- Land consumption
- Bicycle and pedestrian activity and safety
- Bicycle and pedestrian level of service
- Average vehicle occupancy
- Transit productivity (US EPA, 2014)

These measures address transit and active transportation (cycling and walking) use and adoption, which is critical in sustainable transportation and reducing greenhouse gas emissions produced from transportation. Multimodal assessment is also necessary to evaluate the performance of the whole transportation system, since no mode operates in isolation from the others. These are also greatly influenced and largely determined by land-use patterns and other externalities. Ultimately though, these can be distilled to the same lives, time, and money.

The term “Level of Service” is a key part of planning to communicate the impacts of a project on the system, though in 2013 California legislation (SB-743) was signed into law that largely did away with LOS requirements, in favor of people moved (Steinberg, 2013). Lives can crudely be distilled into safety terms of fatalities or injuries incurred while using the transportation system. Often this is a measurement of fatalities of road users, cyclists or pedestrians, or workers (such as highway maintenance or construction crews). Other public health measures, such as the impact on air quality for areas surrounding transportation facilities is also another up-and-coming metric for some transportation projects, though not widely adopted yet. Money as a metric takes on many forms, but also shows how focusing on economic impacts is a pragmatic approach to securing and maintaining funding for transportation programs. Some ways these are measured are by monitoring pavement quality (which has impacts on vehicle wear and emissions, but is also an indicator of pavement performance and potential lifecycle costs), travel times (as a reflection of economic performance, and the external costs of congestion), and other cost benefits achieved through improved operations across all sectors in transportation (from maintenance, to planning, to administration).

There has been considerable work in communicating the value of transportation research to stakeholders that ultimately decide on funding and priorities, but the emphasis has largely been on government research programs. In 2009 the Transportation Research Board (TRB) published the guidebook *Communicating the Value of Transportation Research* (NCHRP Report 610) which gives a broad overview for agencies and research programs to demonstrate their value. It recommends that researchers, “adopt a principle of continual communication as part of [the] research process” (National Cooperative Highway Research Program, 2009). It also stresses that communicating value is context specific, dependent upon the nature of the project, the

desired outcomes, and the intended audience. While the guidebook is clearly intended for research administrators, those involved with technology transfer, and communications teams, their approach of context focused case studies presents a good model. Particularly since many of these research programs have diversity in the scale, scope, and topics of projects, that cut across many of the performance areas described above. The anecdotal case study approach is more feasible in terms of tracking and describing to stakeholders even if it doesn't give a holistic view of a research program. There is also an added level of separation from transportation research programs at academic institutions, since the government agencies largely focus on their research outputs (technical reports) and ignore other research products, such as peer-reviewed articles or data. There is also a disconnect of attribution whereby the time a research project is completed and parts of it are implemented or deployed, frequently the original academic researchers are only mentioned in passing if at all. As a result, connecting academic research to the deployed research can be difficult and labor intensive.

The Institute of Transportation Studies

The Institute of Transportation Studies (ITS) was formed in 1947 by the California State Legislature to perform research to shape the state's transportation needs, and was the first research institute of its kind established in the United States (though several soon followed in other states). As the state and the UC system grew, ITS evolved into a four-campus model with institutes at Berkeley, Davis, Irvine, and UCLA. Each campus performed distinct, and at times complementary (or competing), research projects. ITS funding has been allocated through the Public Transportation Account (PTA) of the state budget, but the total amount given to the campuses remained largely unchanged at $980,000 annually since the 1970s. In 2016, the four campus ITS successfully appealed to Sacramento for a one-time funding increase across the

system to $3 million. In the next fiscal year (2017-2018), continued funding was increased to $5.98 million across the system from revenues in the Road Maintenance and Rehabilitation Account generated from an increase to the state's gas tax (Senate Bill 1, or SB1) (UCOP, 2018). Due to California's peculiar ballot initiative process, this funding wasn't certain until Proposition 6 (which would have repealed the new increase in the gas tax) was defeated in November 2018 ballot (Swan, 2018). Had Prop. 6 passed, most of the UC-ITS research and other efforts would have ceased immediately. Before the funding increases, most of the PTA money was allocated for administrative functions, such as programmatic staffing, research management, and the ITS Library at Berkeley. Funding for the different research centers and groups associated with ITS comes through grants and contracts from a variety of sources: Caltrans, the California Air Resources Board, USDOT, USDOE, EPA, private companies, etc. The extra funding from PTA and now SB1 will support small scale, exploratory research that directly addresses California's transportation needs. These projects will also provide some stable funding and support for graduate students, a common limiting factor in any graduate program.

Through the process of approaching Sacramento for increased funding, the need to demonstrate the value and impact of ITS research across the state was an explicit priority, following similar performance measurement expectations outlined in MAP-21 and the FAST Act. For some projects, such as the integrated corridor management system piloted in Connected Corridors, it was easy to demonstrate the effects of the research on the transportation system, such as a decrease in travel times along the corridors where the system was deployed (California PATH, 2018). Other kinds of research, such as looking at potential policies to encourage the adoption of zero-emission vehicles and effects that might have on greenhouse gas emissions, are much more difficult to quantify. When the four ITS directors went to Sacramento for increased

funding, demonstrating the value of the programs was a necessity. As part of that effort, a comprehensive bibliography of research for each of the ITS campuses was compiled by the ITS librarian, including citations from Google Scholar. This crude method of demonstrating breadth and reach of ITS research illustrated the possibilities of using bibliometrics and other tools to quantify research outputs systematically. These metrics were then used to demonstrate some value of the research programs to stakeholders and demonstrate that the taxpayers money was being used effectively to improve and innovate the state's transportation system. Using this metrics to illustrate the performance of the different ITS groups was both a show of accountability but also promotion -- ITS research is used and respected.

The ITS Library was called upon to compile the bibliometric data after a similar exercise for an external academic program review at ITS Berkeley for the Vice Chancellor of Research in 2015. For that exercise, academic publications (articles, books, chapters, and conference papers) were compiled for all faculty and research staff from 2005 through 2015, which was about 1,000 items in total. Since the focus of that review was on the academic activities and output, publications not in academic venues, such as technical reports and white papers, were excluded. After collecting the publication data, it was evident that comparisons of research productivity across academic researchers in that way provided a limited view. It also must be acknowledged that publication output across different disciplines like the major engineering sub disciplines varies, so directly comparing a civil engineer and an electrical engineer is like apples to oranges (Lillquist & Green, 2010). As an example, the output of the computer science researchers at ITS Berkeley was more prolific yearly than some of the civil engineers, but the civil engineers were more likely to author technical reports for Caltrans or other state agencies which were not included in that citation set. The full collection of publications provided a broad overview of

research activities of ITS Berkeley for the previous decades. It provided a forensic view of successfully completed research for each of the research centers and faculty. Research thrusts, shifts in funding sources, and regular collaborators could be gleaned from looking at the bibliographic information.

The next iterations of the work, such as the bibliometric data compiled for the proposal to Sacramento for the increase in PTA funding included grey literature in the publications set. This was especially important since many of those publications were reports for Caltrans, the California Air Resources Board, and other government agencies. Demonstrating that ITS research meets the needs of California and provides value to the state and others in the transportation community was a critical factor in the increase in PTA and SB1 funding. Though it was obvious that there was room to improve tracking the results, impacts, and potential values of ITS research, how to do so was a lingering question -- one that the research community has been working on for a while.

Tracking and Measuring Impact for ITS

After the initial collection of ITS Berkeley bibliographic data in 2015, the potential to use that data for programmatic assessment was evident. In transportation, like many other disciplines, the allure and impact of simple, straightforward metrics to show the efficacy or dysfunction of a system was powerful. The most well-known example in transportation was the Urban Mobility Scorecard from Texas Transportation Institute (TTI), now the Global Traffic Scorecard from INRIX, which provides clear rankings of traffic congestion in metropolitan areas. So that people could brag that their traffic was indeed the worst. The data and methodologies used to generate the rankings never got as much attention as the simple factors that are frequently cited. Access Across America from the University of Minnesota's Center for

Transportation Studies is another set of measures used to compare different transportation regions that is frequently referenced. Instead of traffic volumes, the focus is on the access to work, services, and housing through different transportation modes. Would it be possible to derive a similar set of metrics for transportation research? No. It would not be possible in any meaningful way because a more qualitative approach would be necessary. There have been attempts to establish some correlation between citations and research performance in transportation, but they rely on the scholarly metrics methods described above which only measure the performance of transportation researchers in a narrowly defined way (purely academic) (Hanssen & Jørgensen, 2014). This narrow approach does not reflect the full potential impacts of any given transportation research project. Other efforts to apply scientometrics to transportation research do not address the inherent limitations of the approach, such as the limited scope to academically published research, but they can be a useful approach to gain insights into network effects and trends in the academic transportation research community (Heilig & Voß, 2015). These techniques can be useful to analyze part of the performance of a transportation research program, but other tactics also need to be employed to include adoption either through practice or policy.

Aside from the fact that it is inappropriate and reductive to compare different kinds of research projects with different outcomes and goals using a simplistic metric, the amount of work to collect the data to even attempt that kind of measurement is onerous and not sustainable. If the publications in question were only from academic venues, then tools like Clarivate or Scopus might be sufficient, but even then, that can be problematic for transportation research which is inherently interdisciplinary and cuts across the typical publishing ecosystems. This is another reason it is compelling to use Google Scholar citations as an indicator despite the

unreliability of their data and uncertainty of methods. Research stakeholders are often pleased that Google Scholar citation numbers tend to be bigger and more impressive, even if the actual value is not explicitly clear. Taking a case studies approach to measuring the impact of research is a compelling idea, but also labor intensive. It requires an ethnographic approach to talking to researchers, stakeholders, and the end users to learn how the research products are consumed and used. Especially since much of this research is freely available online, it's unreasonable to ask anybody who might read a report, "Did you find this useful? How useful?" Looking at citation rates is a somewhat sufficient proxy for this metric, though it just shows the something was regarded enough to be cited, the sentiment and the utility of the citation would still need to be evaluated on a case by case basis. This also assumes the citations are correct enough to be picked up by Google Scholar, which is far from guaranteed, and also requires the research to be cited or mentioned in works that are publicly available to be indexed by Google. For many agencies, technical memos and internal reports are only available to staff on organizational intranet. For some areas of transportation research, where ultimately the real impact of the research comes from adoption or deployment in the field, attribution of that work in the field can be nearly impossible to track. The same is also true for public policy, where a policy recommendation might be written in response to a research project that makes a case for a certain change, but there will not likely be any citation of the research or consultants in the legislation or laws. One proposed framework for assessing the impact of social sector organizations (such as non-profits and nongovernmental organizations), takes a context-sensitive approach that focuses on the scope and scale of an organization's mission and operations (Ebrahim & Rangan, 2014). The framework tracks outputs, outcomes, and impact as they fall on different geographic scales. This

approach could be applied to UC-ITS research and other functions, to make sure the assessment methods match the actual scope and scale of the product.

Documenting PTA/SB1 Projects

When UC-ITS was spun up following the increase in core funding from Sacramento, it was an opportunity to incorporate tracking and measurement of completed research into the entire research process. From the outset, consideration for how deliverables can be improved and organized were built into the research administration process. Research managers were consulted about setting expectations of reporting and assessment to ensure that sentiments the ITS directors and stakeholders in Sacramento were addressed. The importance to being able to demonstrate the reach and impact of these research activities was very clear. This approach builds upon the ideas of accountability and performance measurement that were core to MAP-21 and the FAST Act, which set the agenda for transportation in the US. The ITS Library provided guidance in the initial discussions and planning of the administrative process, stressing the need for unique identifiers to track projects and their deliverables. Incorporating assessment from the inception of a project would generate a richer data set that can lead to more meaningful insights and analysis. A broad overview of the process follows:

1. When projects are approved and vetted by UC-ITS leadership, they are entered into a master spreadsheet and assigned a project number. This number acts essentially as a contract number, which all subsequent research products will be associated with.
2. Relevant project information, such as preliminary investigators, campuses involved, budget and scope of the project, and anticipated deliverables are also included. Not all projects will result in publications. Some will result in meetings, workshops, or trainings, and will not have any final documentation. Other projects might result in an online

resource or tool. For projects that will result in written publications, there is the expectation of at least two publications: a policy brief and a written report.

3. All policy briefs and reports are published on the UC-ITS eScholarship repository, rather than the individual websites of the campus performing the research. Standardized templates are used to give publications a uniform look and guaranteed that all required elements are included. Reports include a technical report documentation page using the FHWA template to present all of the project's information in the document. This is done to show the unified cooperation of UC-ITS but also to ensure there is only one version of the publication to be tracked and cite, making that process as simple as possible.
4. The project number is included on each publication to link it back to any other related works. DOIs are also assigned to each report and publication and policy brief to make citation and attribution even easier. The reports are assigned report numbers that are mostly derived from the project number, but not always. This is done as yet another way to disambiguate reports and make it easier for others to site and reference, following conventions of technical reports from many other agencies, like Caltrans and USDOT. These report numbers are often used for retrieval purposes in other systems and databases that the reports are added to.
5. When the project is completed and closed out, the record is updated. Report numbers, DOIs, and URLs to the reports and policy briefs added to the record. Any projects with outstanding deliverables will be readily apparent.
6. Reports are indexed in TRB's transportation research database TRID, which is widely used by researchers and practitioners. It is also indexed by Google. (Both eScholarship and TRID are also found in Google Scholar searches.)

So far only the initial round of PTA funded research has been completed and published, which has proven to be a good test of the publication process. Even though the technical publications are often an afterthought of the research process, by making it an explicit part of project management the reports and policy briefs are worked into the research process. Making the expectations known from the outset and explicitly including all of the elements that will make tracking feasible reflects a change in culture where there is an understood need for research assessment by all involved. The policy briefs are a new publication format whose focus, as the name suggests, are on policies that could be derived from the research. They are two-page documents that summarize the research in a clear and digestible format, with figures that make the findings easy to understand. Their intended audience are policy makers and advisors in Sacramento, to take some kind of action from the research findings, though they are really for anybody interested in the area who might not have the time (or technical understanding) to read the full report.

Tracking PTS/SB1 Projects Now

With the system described above operational, there is a decent foundation in place to assess any impacts of the UC-ITS research funded through PTA/SB1. While this proposed methodology is intrinsically based upon the situational needs and resources of UC-ITS, it is also meant to be applicable to other research programs. An eventual goal is to have this culture of research assessment adopted by all ITS researchers across the four campuses. These methods could also be adapted to meet organizational needs of other academic research centers whose work intersects policy, practice, and society.

The metrics used to assess the impact of research reflect the priorities of the stakeholders, but also can be sustainably collected given limited resources and time. To this end, the metrics focus

on different areas of potential impact depending on the nature of the research and potential outcomes. This context sensitive approach will focus on the value and impact of projects on an individual basis, measuring the impact of the project against its stated goals. Comparing projects in aggregate will likely happen, but that is not the intended use of these metrics beyond broad statements about the value of transportation research.

Evaluation of individual projects rely on these areas:

- Citations and other references to policy briefs and reports in aggregate as pulled from Google Scholar
- Citations and references to PTA/SB1 funded research in the media and other popular sources
- Academic publications (journal articles or conference papers) derived from PTA/SB1 funded projects, and their scholarly metrics
- Adoption of the research in policy and practice
- Research that builds upon the projects funded by PTA/SB1

Collecting and tracking citations, even though they will likely fall outside the traditional scholarly publishing system, remains the most easily quantifiable metric. This will be done semi-regularly during the year following a project's completion. Assessment of the citation data will not really be useful until 5 years after a project's completion, since that is a reasonable time scale to expect some kind of reception and adoption of the research beyond initial discussions. Using Google Scholar makes the most sense at this time for collecting the citation information because the reports and policy briefs will likely be from technical sources in addition to scholarly sources.

Capturing the mention and discussion of UC-ITS research in the media is also an important component of potential impacts of the research because it elevates the research to the citizens whose tax dollars paid for it. Using web alerts to keep track of media mentions of UC-ITS researchers will be implemented to compliment tracking links of the documents in (either through the URL or through their DOI). This multi-pronged approach is necessary because it is exceedingly common for media articles to discuss research without actually linking to or fully citing the report or article, but writers usually will give attribution to the authors. Using altmetrics to also compile these links would be beneficial, but initial attempts have not been very successful. Preliminary attempts in using altmetric data to describe the reach of some ITS research were not fruitful because grey literature is not really monitored by that system. Tracking academic publications that come out of the PTA/SB1 funded research is another important measure, to show how this research not only contributes back to society but also to the academy. This also accounts for the reality that research takes different formats for different audiences, and to have a holistic view of its impact they all need to be accounted for. Encouraging authors to credit UC-ITS (or SB1) for funding will make this easier to focus tracking efforts, but given the current scale of UC-ITS research staff this can reasonably accomplished by tracking citation for publications beyond those funded by PTA/SB1.

Determining how these research projects are adopted in the field, either through policy, integration in technical documents, or deployment is the most difficult area to measure, but also the most important. The field is where the real, practical impact on society takes place. It could be argued that media attention that changes the attitudes or practices of individuals also has impact in changing human behavior, that is also difficult to track and measure. An ethnographic approach will be needed to fill the gaps that other quantitative and automated methods

(searching, linking, and alerts) will not be able to adequately capture. Keeping track of UC-ITS activities in Sacramento (and other similar venues) will help prioritize what to track. Paying attention to proposed California legislation (such as SB1 and SB743) to identify legislative bills that incorporate policy recommendations from UC-ITS research, will also be necessary. Evaluating the influence of UC-ITS for transportation practitioners will also be largely context specific. Following up with the researchers to see if they are aware of any adoption from their colleagues in the field is one way this information will be collected. Another way will be following developments and changes in relevant technical guidelines either from California or other jurisdictions in California. Most UC-ITS projects have a noted stakeholder associated with them, which will make it easier to target efforts in this area but focus should not be limited to those stakeholders. Good ideas may be adopted by practitioner or agency. And while the focus will be on adoption within California since this research is funded by the state's gas tax, tracking adoption across borders will be important and necessary. Commercialization of research, either through patents, the creation of start-ups, or some other kind of partnership with industry is another area of impact that isn't quite accounted for in this model. This could be folded into tracking policy or other forms of adoption of research products, but the information sources will be quite different as will the time scale.

Including research that builds upon these projects is also an important metric that isn't usually captured by most traditional methods, and looks at a time scale that is not immediate. Recognizing that research projects build upon one another and evolve is an important part of the process. It will be nearly impossible to give full attribution to these projects, particularly since passing instances or ideas might have profound implications for other research in ways not easily attributable. Another area to develop over time is the documenting research project "families",

where related projects naturally build upon one another and are interrelated. Assessing projects in a way that leads to some basic classification will document these relationships in the institute's research portfolio, which will help support giving a longer view of research assessment than would be possible if we were to focus on individual projects alone. The value of an earlier research project might not be apparent until much later when a subsequent project is completed. It also acknowledges that research, change, and progress is a process that builds upon itself. Measures to show the impact of the initial research would include subsequent projects, especially those not funded by SB1.

An assessment portfolio will be created for each project to collate all of these metrics, so that they may be made available for the researchers and UC-ITS staff. This will also be an important mechanism to collect citations, mentions, and other documentation needed for these case studies to be developed over time. As mentioned earlier, while tracking citations and gathering the information should begin shortly after projects are completed, it must continue for years to come gain a meaningful understanding of what impacts may actually come from that research. The projects and their impacts will be measured using criteria based on the performance measures outlined by USDOT and the EPA for transportation. Since there is considerable overlap between those two sets of measures, they can be streamlined a bit. A few other topics, not accounted for in those performance measures that are important to the mission of UC-ITS, such as social equity, will also be included. The proposed measures to track, as makes sense for each project, fit the following thrusts:

- Safety
- Infrastructure condition
- Congestion reduction/Traffic flow

- System reliability
- Freight and goods movement and economic vitality
- Environmental sustainability and impacts
- Reduced project delivery delays
- Mode share
- System performance across modes
- Accessibility
- VMT per Capita
- Land Use and Real Estate
- Transportation Affordability
- Social Equity

By using these categories to evaluate projects, it helps keep like with like. Context sensitive evaluation is important to keep in mind the nature and the scope of the research. It's also important to make sure the outcomes are appropriate for the stated goals of a project. This is to temper any notions of going by raw numbers, in pursuit of the largest, most easily measured numbers that can be generated. They may be impressive but have no real value absent the proper context.

Other Considerations and Potential Next Steps

The outlined methodology above is meant to be practical for organizations of varying sizes to adopt and use. The process of data collection could be improved upon by using systems like CrossReff or ORCID for tracking research outputs. This methodology was also developed for the most part before the November 2018 election, when the UC-ITS funding situation was not entirely certain so potential resources like ORCID for publication aggregation were not

included. Many UC-ITS researchers have ORCIDs since they are mandated in the USDOT Public Access plan, though it is not entirely clear how widely they are used beyond required instances. As the UC-ITS publication system develops and more publications are made with linking DOIs and the like, hopefully CrossRef will be a useful tool in tracking output and citations, though we are not yet there.

Another area that should be folded into this process is data citation and publication. While there has been a lot of discussion and enthusiasm for opening data in transportation, publication of transportation research data is still not very common. One reason for this are the barriers to sharing due to the lack of infrastructure, though that will hopefully change with more stable funding. The main obstacle is the current culture around data in transportation research does not really share or publish the data used in research. Some reasons are due to data agreements and licenses with vendors, or the sensitive nature of the data, but there also isn't really an incentive to otherwise change. As the culture changes, and data is treated as a deliverable and research product like the report, then citation and potential reuse of data generated in UC-ITS research will also need to be accounted for.

Quantifying workforce development is another measure that could be incorporated, but would likely require a very different approach. For projects that fund graduate or undergraduate students, including the how many and which students are supported in the project records would be useful. Students eventually become practitioners in some form, and including this in the assessment of a project makes it a clear relationship to workforce development. For projects that include workshops or some other kind of event, it would be useful to not only note the numbers of attendees but from which sectors they represent. Quantitatively tracking the outcomes of workshops or other public meetings will be difficult, particularly if there is not a publication

accompanying the meeting. Tracking mentions of ITS could help uncover some continuations of ideas from these workshops. Further work and development of performance measures and methods for public engagement fora that goes beyond attendance numbers is needed.

Important as these projects are evaluated is to remind stakeholders that metrics are only part of the evaluative context, and that research outputs cannot be adequately reduced to a simple set of numbers. While that might be eye-catching and help wow the public, a measured approach of powerful anecdotes while collecting a robust data set will be needed. his methodology is designed to be extendable to the entire ITS research portfolio, including projects from other funding sources. Ideally the process could be adapted to work with the individual research centers, so that their administrators can work with the ITS Library in the necessary documentation and preparation at the beginning of the research project to more readily be able to measure and track progress upon completion. That would be a cultural change though and require an infrastructure and continuity that doesn't really exist at this time.

References

A Summary of Highway Provisions - FAST Act | Federal Highway Administration. (n.d.). Retrieved November 28, 2018, from https://www.fhwa.dot.gov/fastact/summary.cfm

Accessibility Observatory at the University of Minnesota. (n.d.). Access Across America. Retrieved November 30, 2018, from http://ao.umn.edu/research/america/index.html

Archambault, É., & Larivière, V. (2009). History of the journal impact factor: Contingencies and consequences. *Scientometrics*, *79*(3), 635–649. https://doi.org/10.1007/s11192-007-2036-x

Auger, C. P. (1998). *Information sources in grey literature* (4th ed). London ; New Providence, N.J: Bowker-Saur.

Banister, D. (2014). Where to Start? *Transport Reviews*, *34*(1), 1–3. https://doi.org/10.1080/01441647.2013.874131

Bornmann, L. (2012). Measuring the societal impact of research: Research is less and less assessed on scientific impact alone—we should aim to quantify the increasingly important contributions of science to society. *EMBO Reports*, *13*(8), 673–676. https://doi.org/10.1038/embor.2012.99

Bornmann, L. (2013). What is societal impact of research and how can it be assessed? a literature survey. *Journal of the American Society for Information Science and Technology*, *64*(2), 217–233. https://doi.org/10.1002/asi.22803

Bornmann, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *Journal of Informetrics*, *8*(4), 895–903. https://doi.org/10.1016/j.joi.2014.09.005

Burawoy, M. (2011). Redefining the Public University : Global and National Contexts. In J. Holmwood (Ed.), *A Manifesto for the Public University* (1st ed., pp. 27–41). London: Bloomsbury Academic. Retrieved from http://www.bloomsburycollections.com/book/a-manifesto-for-the-public-university/ch2-redefining-the-public-university/

California PATH. (n.d.). Connected Corridors Program. Retrieved November 30, 2018, from https://connected-corridors.berkeley.edu/

Ebrahim, A., & Rangan, V. K. (2014). What Impact? A Framework for Measuring the Scale and Scope of Social Performance. *California Management Review*, *56*(3), 118–141. https://doi.org/10.1525/cmr.2014.56.3.118

Editors, T. Pl. M. (2006). The Impact Factor Game. *PLOS Medicine*, *3*(6), e291. https://doi.org/10.1371/journal.pmed.0030291

England, H. F. C. of. (n.d.). What is the REF? - REF 2021. Retrieved November 29, 2018, from https://www.ref.ac.uk/about/what-is-the-ref/

Ernø-Kjølhede, E., & Hansson, F. (2011). Measuring research performance during a changing relationship between science and society. *Research Evaluation*, *20*(2), 131–143. https://doi.org/10.3152/095820211X12941371876544

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2007). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *The FASEB Journal*, *22*(2), 338–342. https://doi.org/10.1096/fj.07-9492LSF

Haddaway, N. R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching. *PLOS ONE*, *10*(9), e0138237. https://doi.org/10.1371/journal.pone.0138237

Hanssen, T.-E. S., & Jørgensen, F. (2014). Citation counts in transportation research. *European Transport Research Review*, *6*(2), 205–212. https://doi.org/10.1007/s12544-013-0122-0

Heilig, L., & Voß, S. (2015). A Scientometric Analysis of Public Transport Research. *Journal of Public Transportation*, *18*(2). https://doi.org/10.5038/2375-0901.18.2.8

INRIX. (n.d.). INRIX Global Traffic Scorecard. Retrieved November 30, 2018, from http://inrix.com/scorecard/

Institute of Transportation Studies. (n.d.). Retrieved from https://escholarship.org/uc/ucits

Kochetkov, D. M. (2018). A Correlation Analysis of Normalized Indicators of Citation. *Publications*, *6*(3), 39. https://doi.org/10.3390/publications6030039

Korevaar, J. C. (1996). Validation of bibliometric indicators in the field of mathematics. *Scientometrics*, *37*(1), 117–130. https://doi.org/10.1007/BF02093488

Lawrence, A., Houghton, J. W., Thomas, J., & Weldon, P. (2014). Where is the evidence: realising the value of grey literature for public policy and practice. Retrieved from http://apo.org.au/node/42299

Lillquist, E., & Green, S. (2010). The discipline dependence of citation statistics. *Scientometrics*, *84*(3), 749–762. https://doi.org/10.1007/s11192-010-0162-3

Manville, C., Guthrie, S., Henham, M.-L., Garrod, B., Sousa, S., Kirtley, A., … Ling, T. (2015). Assessing impact submissions for REF2014 [Product Page]. Retrieved November 27, 2018, from https://www.rand.org/pubs/research_reports/RR1032.html

MAP-21 - Fact Sheets - Performance Management | Federal Highway Administration. (n.d.). Retrieved November 28, 2018, from https://www.fhwa.dot.gov/map21/factsheets/pm.cfm

National Cooperative Highway Research Program, Transportation Research Board, & National Academies of Sciences, Engineering, and Medicine. (2009). *Communicating the Value of Research: Contractor's Final Report*. Washington, D.C.: Transportation Research Board. https://doi.org/10.17226/23034

National Institutes of Health. (n.d.-a). Federal RePORTER - Smart Search: Find federal agencies scientific awards data from this easy to use seamless search interface. Retrieved January 3, 2019, from https://federalreporter.nih.gov/

National Institutes of Health. (n.d.-b). STAR METRICS® - Home. Retrieved January 3, 2019, from https://www.starmetrics.nih.gov/

Nightingale, P., & Scott, A. (2007). Peer review and the relevance gap: Ten suggestions for policy-makers. *Science and Public Policy*, *34*(8), 543–553. https://doi.org/10.3152/030234207X254396

NSF - National Science Foundation. (n.d.). STAR METRICS: New Way to Measure the Impact of Federally Funded Research. Retrieved January 3, 2019, from https://www.nsf.gov/news/news_summ.jsp?cntn_id=117042

Official DOT Public Access Plan v1.1. (2015, December 18). [Text]. Retrieved November 30, 2018, from https://www.transportation.gov/mission/open/official-dot-public-access-plan-v11

Parker, J., & Teijlingen, E. van. (2012). The Research Excellence Framework (REF): Assessing the Impact of Social Work Research on Society. *Practice*, *24*(1), 41–52. https://doi.org/10.1080/09503153.2011.647682

Research Publication Development Tools | FHWA. (n.d.). Retrieved November 28, 2018, from https://highways.dot.gov/research/resources/research-publication-development-tools

Salter, A. J., & Martin, B. R. (2001). The economic benefits of publicly funded basic research: a critical review. *Research Policy*, *30*(3), 509–532. https://doi.org/10.1016/S0048-7333(00)00091-3

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ*, *314*(7079), 497. https://doi.org/10.1136/bmj.314.7079.497

Shotton, D. (2013). Publishing: Open citations. *Nature News*, *502*(7471), 295. https://doi.org/10.1038/502295a

Steinberg. Environmental quality: transit oriented infill projects, judicial review streamlining for environmental leadership development projects, and entertainment and sports center in the City of Sacramento., Pub. L. No. SB-743. Retrieved from https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201320140SB743

Swan, R. (2018, November 7). Prop. 6: Californians reject gas tax repeal measure - SFChronicle.com. *San Francisco Chronicle*. Retrieved from

https://www.sfchronicle.com/politics/article/Prop-6-California-gas-tax-increase-holds-narrow-13369294.php

Taubes, G. (1993). Measure for measure in science. *Science*, *260*(5110), 884–886. https://doi.org/10.1126/science.8493516

Terämä, E., Smallman, M., Lock, S. J., Johnson, C., & Austwick, M. Z. (2016). Beyond Academia – Interrogating Research Impact in the Research Excellence Framework. *PLOS ONE*, *11*(12), e0168533. https://doi.org/10.1371/journal.pone.0168533

Texas Transportation Institute. (n.d.). Urban Mobility Scorecard — Urban Mobility Information. Retrieved November 28, 2018, from https://mobility.tamu.edu/ums/

UC Institute of Transportation Studies (ITS) | UCOP. (n.d.). Retrieved November 26, 2018, from https://www.ucop.edu/research-graduate-studies/programs-and-initiatives/state-and-systemwide-research/transportation-initiative.html

United States., & Bush, V. (1945). *Science, the endless frontier. A report to the President*. Washington: U.S. Govt. print. off. Retrieved from https://catalog.hathitrust.org/Record/001474927

US EPA, O. (2014, January 7). Guide to Sustainable Transportation Performance Measures [Reports and Assessments]. Retrieved November 29, 2018, from https://www.epa.gov/smartgrowth/guide-sustainable-transportation-performance-measures

Wagner, C. S., & Alexander, J. (2013). Evaluating transformative research programmes: A case study of the NSF Small Grants for Exploratory Research programme. *Research Evaluation*, *22*(3), 187–197. https://doi.org/10.1093/reseval/rvt006

Watermeyer, R. (2016). Impact in the REF: issues and obstacles. *Studies in Higher Education*, *41*(2), 199–214. https://doi.org/10.1080/03075079.2014.915303

Watermeyer, R., & Hedgecoe, A. (2016). Selling ‘impact’: peer reviewer projections of what is needed and what counts in REF impact case studies. A retrospective analysis. *Journal of Education Policy*, *31*(5), 651–665. https://doi.org/10.1080/02680939.2016.1170885