









ARTICLE



<https://doi.org/10.1057/s41599-020-00647-z>

OPEN

The transformative power of values-enacted scholarship

Nicky Agate ¹, Rebecca Kennison ², Stacy Konkiel³, Christopher P. Long ⁴✉, Jason Rhody ⁵, Simone Sacchi ⁶ & Penelope Weber ⁵

The current mechanisms by which scholars and their work are evaluated across higher education are unsustainable and, we argue, increasingly corrosive. Relying on a limited set of proxy measures, current systems of evaluation fail to recognize and reward the many dependencies upon which a healthy scholarly ecosystem relies. Drawing on the work of the HuMetricsHSS Initiative, this essay argues that by aligning values with practices, recognizing the vital processes that enrich the work produced, and grounding our indicators of quality in the degree to which we in the academy live up to the values for which we advocate, a values-enacted approach to research production and evaluation has the capacity to reshape the culture of higher education.

¹University of Pennsylvania Libraries, Philadelphia, PA 19104, USA. ²KN Consultants, New York, NY 10025, USA. ³Altmetric, Baltimore, USA. ⁴Michigan State University, East Lansing, MI 48824, USA. ⁵Social Science Research Council, Brooklyn, NY 11201, USA. ⁶European University Institute, 50014 Fiesole, Italy. ✉email: cplong@msu.edu

Introduction

“The tyranny of the quantifiable is partly the failure of language and discourse to describe more complex, subtle, and fluid phenomena, as well as the failure of those who shape opinions and make decisions to understand and value these slipperier things. It is difficult, sometimes even impossible, to value what cannot be named or described, and so the task of naming and describing is an essential one in any revolt against the status quo of capitalism and consumerism”.

— Rebecca Solnit (2014, p. 97)

The current mechanisms by which scholars and their work are evaluated across higher education are unsustainable and, we argue, increasingly corrosive. Relying on a limited set of proxy measures, current systems of evaluation fail to recognize and reward the many dependencies upon which a healthy scholarly ecosystem relies.¹ The publication of a journal article or a monograph, for example, depends upon feedback from reviewers whose insights and effort enhance the scholarship while they and their work remain largely invisible, unacknowledged, and undervalued. If, as Bergeron et al. (2014) have argued, “research productivity is the determining factor in academic rewards” but the conditions of scholarly production are ultimately unsustainable, the mechanisms of scholarly evaluation are caught up in a pernicious cycle in which the academy rewards products of scholarship without acknowledging and sustaining the very processes that enhance their production. Worse still, traditional bibliometrics and alternative metrics (“altmetrics”) that attempt to designate quality and measure “impact” in strictly quantifiable terms have a flattening and alienating effect. By assigning a score or set of scores—like the number of citations, mentions, or shares—as proxies of quality, such metrics are unable to effectively account for nuances of context, depth of engagement, integrity of process, or sophistication of argument. Metrics flatten out these aspects of scholarship that testify to its quality. They seduce scholars into caring more about these proxies than the quality and impact they purport to indicate. This obsession with a limited number of proxy measures devalues a wider diversity of scholarly contributions, such as public-oriented work, and erases the labor of those outside traditional scholarly pathways. Current engines of evaluation not only mistake quantity of production for quality of scholarship, they also fail to recognize and sustain the many scholarly activities not captured by these mechanisms that nevertheless enrich academic and public life.

While the contextual frame for the current essay is largely based in the United States, we have engaged with like-minded initiatives within the European context. Though the framing here may be constrained by nuances germane to the North American context, particularly as it relates to tenure system processes, we suggest that these interventions are applicable to higher education systems across the globe not least because to be effective they must be situated within and adapted for local contexts.

As more provosts demand dashboards of quantitative metrics to evaluate the impact of faculty scholarship and as more humanities and social sciences (HSS) promotion-and-tenure dossiers and faculty annual reviews are required to include citation counts and other so-called research impact metrics, there is a growing sense among scholars that they are being evaluated on what is easily measured rather than on any holistic understanding of the impact their teaching, service, and scholarship actually has. Such attempts to determine quality by measuring quantity, using a toolset that captures a shockingly narrow set of outputs when considered against the many activities that comprise a scholar’s

practices, fuel a toxic culture predicated on scarcity, competition, and alienation from personal and institutional values.²

Reversing this insidious cycle of alienation requires interventions at three levels, the development of which have formed the core of the work of the authors in their roles as co-PIs on the Humane Metrics in Humanities and Social Sciences (HuMetricsHSS) initiative. The first intervention calls for a realignment of how the academy conceptualizes and assesses academic practice, shifting from a focus on research output to one on scholarly process. The second insists that when research products are broken down into their constituent processes—the how rather than the what—what is reinforced is (a) the primacy of the end product as the ne plus ultra of scholarship and (b) the myth of the lone scholar. Privileging the product elides the diverse amalgam of activities that contribute to the quality of the outcome. Critical to the success of a digital humanities project, for example, are activities ranging from data curation to coding to project management to accessibility review; for a research article, to take another example, critical activities include peer review, editing, and proofreading, the intellectual work for which is barely (if ever) recognized in product-oriented evaluation metrics. The myth of the lone scholar similarly fails to recognize a wider cast of actors who contribute substantively to the quality of the scholarship produced—the librarians and community activists and editors and mentors who give the scholarship texture and depth. The third and most fundamental intervention in our work is a focus on engaging institutions to establish contextual sets of carefully negotiated values that are agreed upon by their scholarly communities and enacted through scholarly processes and practices. By leading scholars at a given institution or in a given department, school, or center through structured conversations designed to reconnect them with the core values that animate their professional work and their institution’s mission, we seek to cultivate fulfilling habits of scholarship and provide a framework for more meaningful modes of evaluation. Aligning academic recognition and reward with such a values framework has a powerful transformative effect. It allows for the acknowledgment of the vital and often undervalued processes and people that enrich scholarship and scholarly life, and it can nudge institutions toward the fulfillment of the promise of their mission statements or strategic plans in more than name alone. In fact, we would argue that for scholars, scholarly work, or scholarly processes to be deemed “excellent”³—or for institutions to be highly ranked or library collections to be deemed of “high quality”—“excellence” and “quality”, which too often serve as empty signifiers, must be recast in terms of how effectively an explicit set of shared and agreed-upon values are instantiated in the work(s) produced (For more angles on this argument, see Kraemer-Mbula et al. (2020)).

The HuMetricsHSS initiative, together with other efforts that are developing complementary frameworks and approaches, builds on the three levels of interventions described above to stimulate and support a transformative process that, we argue, is capable of reshaping the culture of higher education, starting with the humanities and social sciences.

Perverse incentives

Evaluation systems and the activities they incentivize are particularly ill-matched to the values that animate much research in HSS disciplines. Current incentive structures not only do not encourage but in fact often discourage scholars from engaging in work—from formative review to mentoring, from community-engaged participatory research to collaborative interdisciplinary scholarship—that is necessary to maintain flourishing communities in which research and pedagogy can thrive. These practices

nevertheless are essential pillars of the infrastructure of the academy. They deepen our understanding of the issues we study, enhance our relationships with colleagues, and actively enrich public life.

Applications of metrics in HSS contexts, especially against the backdrop of efforts like the Research Excellence Framework (REF) in the United Kingdom, have generated considerable skepticism. Such evaluation approaches neither correctly assess the quality and impact of HSS scholarship nor serve the best interest of HSS researchers. Much HSS research requires lengthy time to publication, significant developmental editing in the publication process, and a broad ecosystem of scholarly practices and objects (not least of which is the most-rewarded measure for the purposes of promotion and tenure in the humanities: the scholarly monograph). Even the fairly recent inclusion of monographs in commercial bibliometrics databases such as Scopus and the existence of Clarivate Analytics' Book Citation Index does not counteract the fact that the scores assigned to researchers for "impact" (the "h-index", for instance) are unevenly weighted against traditionally single-author monographic disciplines. Compared with their STEM colleagues, researchers in HSS cannot produce enough or publish fast enough or get cited quickly enough for their work to seem to count as impactfully important. To make matters worse, as noted by Simon Baker (2018), much of this work is not cited at all.

Jeanette Hatherhill (2018) has argued that many scholars in HSS fields have by and large rejected current metrics systems because of their "exclusion of non-traditional scholarship and interdisciplinary research, difficulty in sub disciplinary application, replication of systemic gender bias, and complicity in the corporatization of the academy". They are inherently biased toward fast-paced, highly competitive, and article-focused disciplines.

Even altmetrics, skeptics argue, cannot convey the breadth of HSS research engagement. As shown by Bornmann (2014) and Scott (2012), social media attention does not equal quality, relevance, or impact, and alternative metrics cannot account for more ephemeral, yet essential, interactions such as classroom debates, public discussions, or conversations at conferences—nor can any metrics account for the full range of scholarly activities performed by contingent and tenure-system faculty alike, whether they work in a community or small liberal arts college, a comprehensive regional institution, or a national research-intensive university. Furthermore, the metrics systems currently adopted (and abused) in academic evaluation are designed for and mostly focus on summative assessments of the final output of research and not on the kind of formative assessments that would help researchers and scholars improve their practices and elevate the quality of their scholarship. No wonder, then, that researchers such as Hicks et al. (2015) and Laudel and Gläser (2006) or those at even research-intensive universities such as Rutgers (Flaherty, 2016) have grave concerns regarding the potential uses and abuses of metrics in personnel reviews. By establishing incentives for advancement that focus on canonical or well-recognized indicators, such as citations to and in journal articles, such metrics distort the reality of academic labor and undermine a more holistic approach to scholarly life.

Given the perverse incentives enabled by current mechanisms of evaluation in higher education, is it any surprise that scholars act in self-serving ways: they treat graduate students with disdain (K. J. Baker, 2018; Braxton et al., 2011; Noy and Ray, 2012), steal each other's ideas (Bouville, 2008; Douglas, 1992; Grossberg, 2004; Hansson, 2008; Lawrence, 2002; Martin, 1997; Resnik, 2012), engage in citation gaming practices (Baccini et al., 2019; Cronin, 2014; Gruber, 2014; Rouhi, 2017; Sabaratnam and Kirby,

2014) such as "citation cartels" (Franck, 1999; Onbekend et al., 2016) or even outright citation malpractice (Davenport and Snyder, 1995), cite only those with whom they agree (Hojat et al., 2003; Mahoney, 1977), insist that their Ph.D. students cite them in every work (Hüppauf, 2018; Sugimoto, 2014), require undergraduates buy their \$200 book,⁴ manipulate images to better suit their argument (Clark, 2013; Cromey, 2010; Jordan, 2014), manipulate *p*-values (Gelman and Loken, 2013; Head et al., 2015; Wicherts et al., 2016),⁵ denigrate competitors' research in peer review (Baliotti et al., 2016; Lee and Schunn, 2011; Mahoney, 1977; Mallard et al., 2009; Penders, 2018; Rouhi, 2017)—or openly ridicule earnest peer review of what turn out to be hoax papers (Mounk et al., 2018; Schliesser, 2018; White, 2004)—or change their research to suit the metrics, as Aagaard et al. (2015) and Díaz-Faes et al. (2016) and many others (Chubb and Reed, 2018; Gingras, 2016; Hammarfelt and de Rijcke, 2015; Pontille and Torny, 2010; Rafols et al., 2012) suggest often happens? These symptoms of a toxic culture are shaped and sustained by a neoliberal ideology that understands education as a private good oriented toward idiosyncratic ends within an environment of heightened competition (Fitzpatrick, 2019, pp. 26–33).

Toxic culture. In his speculative video *Our Program* (2016), Zachary Kaiser follows the trajectory of this neoliberal ideology to a logical, dystopian conclusion. In so doing, Kaiser uncovers three dimensions of the ideology (mis)shaping higher education that need to be redressed if a more humane, enriching, and supportive academic culture is to be nurtured and sustained. *Our Program* imagines a not-too-distant future in which small stock market-like tickers are mounted outside of the office of every faculty member to track in real time and to showcase for everyone the evaluative metrics of each member of the faculty—their so-called Faculty Productivity Index (FPI). The effect of the appearance of these tickers on the imagined department is to make palpable the competitive arena in which the faculty are situated, to exert relentless time pressure upon them to be efficient and productive, and to further alienate them from the work they value most. As the disembodied voice of the video says of the presence of the tickers: "It became part of our physical and psychic landscape—our numbers always there for us and our colleagues to see". Even if there are not (yet) physical tickers installed outside of every faculty office across higher education, the obsession with metrics has already distorted the psychic landscape of the academy. The contours of this distorted landscape are determined, we argue, by three primary dimensions: the pervasive culture of competition, the accelerated rhythms of academic time, and the alienation from core values.

Higher education has come to be dominated by a concern with individual distinction that drives constant competition within a prestige economy. In *Generous Thinking* Kathleen Fitzpatrick captures the manner in which competition saturates our thinking and subverts our attempts to collaborate when she writes:

Always, always, in the hidden unconscious of the profession, there is this competition: for positions, for people, for resources, for acclaim. And the drive to compete that this mode of being instills in us can't ever be fully contained by these specific processes; it bleeds out into all areas of the ways we work, even when we are working together. (Fitzpatrick, 2019, pp. 26–27)

This pervasiveness of competition (in the face of limited resources, whether real or imagined) undermines scholarly collaboration and the shared creation of knowledge by infusing our work together with an intense obsession with individual distinction that privileges corrosive forms of interaction. These

competitive forms of interaction have become internalized in such a way that scholarship itself has become identified with a toxic agonism and exclusionary practices that threaten its very existence. Fitzpatrick puts it this way:

That need to be competitive, as it manifests within the university, leads scholars to adopt pugilistic forms of critique, as well as styles of discourse that exclude the uninitiated, as the primary modes of engagement with our work. These modes allow us not only to demonstrate our dominance over the materials that we study and the ways that those who've gone before us have studied them, but also to establish and maintain our standing within the academic marketplace. (Fitzpatrick, 2019, pp. 129–130)

The drive to establish and then maintain standing within the academic marketplace also distorts the rhythm of academic time. Scholarly work is disciplined by the temporality of the tenure clock and the accounting practices associated with the metrics of production that allegedly determine prestige.⁶ By parsing scholarly activity into discrete categories of teaching, research, and service, the contemporary tenure and promotion process confuses means with ends, segregates integrated modes of scholarly engagement, and imposes reified time constraints on essential scholarly processes that when done well—and with care for the work and for one another—take more time. As Mountz et al. (2015) put it, “the overzealous production of research for audit damages the production of research that actually makes a difference”.

This distorted relationship with time has deleterious effects on scholarship and on the scholarly community. With regard to the former, neoliberal institutional time defers the possibility of being present to the unfolding of an idea, disciplines the pursuit of a thread of inquiry that may lead in unanticipated directions, and truncates the advancement of knowledge by flattening our attempts to discover depth of meaning. In reflecting on her experience working with St. Joseph's Community School in the early 1970's, Patricia Hill Collins emphasizes the generative power of stepping outside of the debilitating structures imposed by higher education:

St. Joseph enabled me to explore the connections among critical pedagogy, engaged scholarship, and the politics of knowledge production, delaying for a decade the deadening ‘publish or perish’ ethos of higher education. Instead, it put me on a different path of being a rigorous scholar and public intellectual with an eye toward social justice. (Collins, 2012, p. 17)

Escaping the deadening ethos of higher education enabled Collins, “to experience ideas and actions as iterative” (p. 16). Collins' experience suggests too the degree to which a distorted relationship with time undermines attempts to integrate iterative and intersectional practices of scholarship—ones where the process matters as much as the output—into the life of the university in ways that affirm the work and experiences of traditionally underrepresented minorities. Such practices take time, and they require care in relation to both the people and the issues involved. Careful and caring scholarship requires a level of attention that cannot easily be accomplished within the “deadening ‘publish or perish’ ethos of higher education”, for, as Virginia Held (2015) has emphasized, in an ethics of care “one attends with sensitivity to particular others in actual historical circumstances, one seeks a satisfactory relation between oneself and these others, one cultivates trust, one responds to needs, aiming at and bringing about as best one can the well-being of the others along with that of oneself”.

If the culture of competition and the acceleration of time characterize two dimensions of the distorted landscape of contemporary higher education, an alienation from values names an extremely pernicious third.

Indeed, one of the greatest ironies of the scarcity-fueled system of hypercompetition in which scholars in HSS now find themselves is the apparent mismatch between scholars' own values and those they believe the academy to hold. A recent survey by the ScholCommLab (Niles et al., 2020) that asked faculty members why they publish where they do elucidates this phenomenon. “Put plainly”, the authors write, “our work suggests that faculty are guided by a perception that their peers are more driven by journal prestige, journal metrics (i.e., JIF [journal impact factor] and journal citations), and money (i.e., merit pay) than they are, while they themselves value readership and open access of a journal more”. The study suggests that junior faculty tend to adjust their behavior to fit not their own values, but rather those they perceive to be held by their senior colleagues (who, the team's research finds, are actually even more likely to value readership, openness, and other non-traditional indicators of prestige). The researchers extrapolate their findings to the tenure and promotion process, where, they argue, junior and non-tenured faculty appear to believe that those evaluating them care about publication frequency and publication in high-ranking journals above all else, while the tenured faculty they interviewed—those who serve on tenure and promotion committees—typically take a more nuanced view. There appears to be a curious slippage occurring between perceived and actual values that ends up disempowering junior scholars by leading them to view the processes or standards by which they are judged not only as the inalterable dictates of an alienated and depersonalized administrative body, but also as the very structures that they must replicate and reinforce as they become more senior in order to legitimize their status (Diamantes, 2005; McKiernan et al., 2019; Zaccchia, 2017). The slippage occurring between perceived and actual values manifests itself even more so when those academics who find themselves involved in evaluation processes seem forced by current evaluation systems and established behaviors to neglect (or forget) the values they themselves claim as their own. Even if frank conversations about possible change occur, there is a regressive dimension of evaluation that regularly returns to a common refrain of “this is how it's always been done” or an assertion that the existing process is the only means of establishing objectivity at scale while remaining competitive in a prestige economy (For more on this point, see, e.g., Moher et al., 2018).

An intervention that has sought to redress the regressive dimension of traditional evaluation processes at scale is the Cultivating Pathways of Intellectual Leadership (CPIL) initiative of the College of Arts and Letters at Michigan State University (MSU). Leveraging the existing processes of the promotion and tenure process at a research-intensive land-grant university (https://en.wikipedia.org/wiki/Land-grant_university), the CPIL framework is designed to empower scholars to identify career goals toward which their work is oriented, including a consideration of the values that will inform the scholarship they undertake. By reorienting the existing cadence of formative mentoring conversations between scholars and their supervisors toward an explicit articulation and documentation of the goals, values, and practices that will shape the work in the short, medium, and long term, the CPIL approach seeks to intentionally align personal and institutional values such that the practices and products of scholarship enact the agreed-upon values. This approach does not require a radical departure from the practices

that currently shape the promotion-and-tenure process at major research universities. It does, however, require a subtle but important shift in thinking about the traditional three legs of the tenure-and-promotion stool (i.e., research, teaching, and service). A discussion of this model by the MSU College of Arts and Letters leadership team puts it this way:

By empowering colleagues to identify the core values that animate their work, and providing them with a structure that can value this work in a variety of forms, we've had to reimagine the tenure and promotion process in the College by shifting its focus from means (teaching, research, service) to ends (sharing knowledge, expanding opportunity, mentorship and stewardship). This shift opens new opportunities to recognize and reward a wider variety of activities as contributing to the core mission of the university. (Cilano et al., 2020)

This shift from means to ends enables those scholars who understand their teaching, research, and service work as intimately woven together to tell a more textured story about the impact of their work. The CPIL framework thus challenges traditional dichotomies between activism and scholarship, teaching and research, and individual and collaborative accomplishment in ways that enable institutions to recognize and reward a wider diversity of scholarship and support faculty committed to undertaking meaningful work that enacts the values about which they care most deeply.

Aligning values with practices requires, as Fitzpatrick has argued, “transforming the thinking within our institutions in ways that will enable us to argue in new terms, on our own terms, for the real nonmarket, public service function of our institutions of higher education with those outside of them” (2019, p. 195). Such transformative thinking challenges us not only to trace the source of toxicity to the neoliberal ideology that instrumentalizes education and shrinks it to a private good for autonomous individuals, but to articulate and advocate for ways to measure up to an ideal of education as a public good oriented toward creating more just realities within complex, intersectional, and interdependent communities.

Even if the academy is beginning to turn in this direction, the great distance we have yet to travel is clear from the following two lived experiences.

In 2018, Yvette DeChavez, then a postdoctoral lecturer in American literature, designed an “inherently political” syllabus and course description, in which the authors taught included no white men. A professor in the department where she was working suggested not only that she adapt it to include “more canonical (i.e., white male)” writers, but that it would be prudent to do so if she wanted to have appeal on the job market. She did not make the change. Instead, as she writes in an *Los Angeles Times* article, “My students not only appreciated it, they grew from the experience. On more than one occasion, a student of color broke down in tears, expressing that it was the first time they'd ever read something by someone like them in a university setting” (DeChavez, 2018). If that's not what we mean when we talk about impact, we're doing it wrong.

DeChavez, a successful researcher and writer focusing on “centering the voices, narratives, and performances of Indigenous Americans and People of Color in academia and the media”, now bills herself as a “recovering academic” (yvettedechavez.com). Her recovery is academia's loss. Here is a first-generation scholar who actively intervened in the colonialist structures that underpin academe, who lived and adhered to her values despite being told that doing so would make her less valuable in the academic ecosystem, and whose experiences in that ecosystem led her to abandon it altogether.

If DeChavez's experience speaks to what leads scholars to leave the academy, Michaela McSweeney's experience highlights the price so many pay to enter it. In an interview with Skye Cleary on the American Philosophical Association blog, McSweeney, now an assistant professor of philosophy at Boston University, speaks candidly about her experience in graduate school: “I felt like no one understood me and like I was completely out of place, an alien. I also felt like a traitor to my own values, a lot of the time. Maybe almost all of the time”. She goes on to provide this advice to other scholars:

[D]on't let people make you believe that you have to make compromises about whatever it is that matters to you the most, about your deepest moral and personal commitments. You don't. Nothing is worth doing that for; if you have values that you must completely and continuously compromise in order to “be successful” in something, then that job and that success will not make you happy. This should be obvious to us, but I think academic careers suck you in so much that somehow it is not obvious in this context. (Cleary, 2018)

McSweeney names here the distorting effect of an academic context that has become toxic. The metrics we have historically used to quantify “success” in higher education are not rooted in shared values that the higher education endeavor purports to uphold. As the tickers in Kaiser's speculative imaginary so eloquently demonstrate, our tendency to value what can be counted alienates us from ourselves and others and prevents us from undertaking meaningful work that might have transformative impact. A change in orientation is required if we are to redress the toxic culture of higher education.

Just such a change is suggested by Moore et al. in “Excellence R Us: University Research and the Fetishisation of Excellence” (2016), who propose that the rhetoric of “excellence” be substituted with that of “soundness” and “capacity building”, to address the hypercompetitive *modus operandi* that is so pervasive in current practices of research and scholarship. They claim that “the evaluation of ‘soundness’ is based on the practice of scholarship, whereas ‘excellence’ is a characteristic of its objects (outputs and actors)” (2016, p. 8). Indeed, scholarly production is iterative, has interwoven dependencies, and is contextualized by community, all of which suggests the need for a more sophisticated assessment of process rather than a crude accounting of product. “In this sense”, as Moore et al. argue, “soundness aligns well with approaches that locate the value of scholarship and evaluation in the nature of its processes (that is, ‘proper practice’) and its social conduct” (2016, p. 8). While the shift from “excellence” to “soundness” goes some distance in redressing a toxic culture of competition by directing attention to structures of argument, this rhetorical move remains too limited in scope—it redresses just one among a broader array of values.

The rhetoric of excellence has created a culture of hypertoxicity and competition, and, as we argue in our introduction, a corrective shift from an evaluation of outputs to one of processes is a crucial step toward cultivating a healthier academic ecosystem. A focus on processes and practices lays bare much of the undervalued and largely invisible labor that is both critical to enhancing the quality of the work produced and constitutive of a deeper, more fulfilling sense of what Bruce Macfarlane has called “academic citizenship” (2007, 261). The wide range of activities associated with academic citizenship—from peer reviewing to mentoring, from teaching observations to leadership in scholarly societies or on university committees—are key to the success of any academic endeavor. However, as Kathy Lund Dean and Jeanie M. Forray (2018) argue about peer review, we currently operate in “a performance evaluation system that

rewards publications as academic scholarship while rendering invisible a fundamental component in the production of that scholarship”). If we were to break down the output that is a monograph, for example, into the processes and the people on which it depends, we would quickly see the false narrative of the single-author product dissolve into a rich array of vital interconnected processes performed by myriad people, ranging from the curation of bibliographic material to the formative and evaluative review provided by other scholars.

In the current system, we overemphasize and disproportionately reward scholarly output as a lone endeavor, paying little to no attention to the contributions many colleagues provide, and thereby starving the roots of an ecosystem in favor of a single tree. Our forms of measurement and recognition distort and destabilize this ecosystem, impoverishing the very scholarly practices upon which it depends. Without a systemic appreciation and recognition of the work of peer review, for example, can we really be surprised that editors are having “increasing difficulty” (Fernandez-Llimos, 2019, p. 1502) and “challenges” (Dean and Forray, 2018, p. 166) or just a plain old “tough time” (Kulkarni, 2016) recruiting peer reviewers? Just as the goal of the peer review process is to improve scholarship (even if, given its many detractors, such goals are arguably met), its focus on process—on the *doing* of scholarly work—opens up a pathway toward redefining the quality scholarship and the many practices that enrich its production.

The intentional turn to values

Our argument thus far suggests that current metrics undermine the ecosystem created to generate high-quality scholarship and fail to support and enrich the breadth and depth of scholarly practice that in fact does elevate the quality of scholarship. The current mechanisms of evaluation perpetuate and operate with an impoverished definition of “scholarship”, one that is limited to a narrow set of products (largely monographs and articles) that are counted as artifacts of scholarship in isolation from the broad array of processes and practices that contribute to their creation and enhance their quality. Fetishizing products over processes in this way not only undermines the vitality of the labor that makes scholarship of the highest quality possible, but also undervalues the many laborers—beyond individual authors—whose talent, wisdom, time, and insights enrich the practice of scholarship itself. Perpetuating the myth of the solitary scholar reinforces a values system predicated on the empty rhetoric of “excellence” critiqued by Moore et al. (2016)—one that is overly reliant on proxy indicators to convey the status or success of a final object. Further, while those indicators, as the results of Niles et al.’s recent research suggest, may reflect what scholars *perceive* to be most valued by their peers, the opposite may in fact be the case, creating an Abilene Paradox that threatens to undermine the ecosystem upon which robust scholarship relies. These conditions—perverse incentives, toxic environments, misaligned and misperceived values, coarse proxy measures, and the undervalued and unrecognized breadth of labor and laborers—increasingly compromise the full and fecund potential of scholarship.

If the perceptions of what “excellence” and “quality” actually *are* and what others *think* they are have become uprooted, a first step in preparing the ground for new and deeper roots of scholarly excellence is to realign the systems of evaluation with an explicit understanding of the values that shape scholarship of the highest quality. In valuing what can be easily measured, metrics distort what we most deeply value.

Instead of valuing what can be measured, as we so often do now, we must find ways to measure what we deeply—and sometimes unconsciously—value. Unless scholars work to

actively listen to one another and engage in meaningful conversations about their shared (or conflicting) values, the prestige machine on which the academy runs will continue to reproduce itself, even when that is perhaps counter to what many involved actually want. As the ScholCommLab researchers argue, “If faculty truly value journal metrics and prestige outcomes less than readership and peers reading their work, but perceive ‘others’ to be the promoters of these concepts, fostering conversations and other activities that allow faculty to make their values known may be critical to addressing the disconnect. Doing so may enable faculty to make publication decisions that are consistent with their own values”. (Niles et al., 2020, p. 9)

Structured conversations about the values that animate the most innovative and meaningful scholarship have transformative power because they enable us to align the values about which we care most deeply with practices that produce scholarship of the highest quality. To redress the challenges we have articulated above, values-alignment needs to be recognized as a catalyst of scholarly excellence.

Often, the academy uses prestige—the imprimatur of established organizations such as publishers, aspirations toward “peer institutions”, and other proxies—to signify excellence or quality. Rather than focusing on outputs and assigning them a similarly isolated quantitative proxy of “excellence”, the HuMetricsHSS approach proposes a realignment of focus to the many *processes* of scholarship—both because such a realignment offers an opportunity to broaden the spectrum of what “counts” and because breaking down a product into its constitutive processes enables us to recognize and reward practices of scholarship that enhance the quality of the work produced. For example, as mentioned, the vital work of peer review remains largely invisible despite its critical role in enhancing scholarship through the creative insights, thoughtful feedback, and intellectual effort of expert colleagues. By counting only the product of such rich scholarly engagement as the primary measurable outcome of the scholarship of a single author or small group of collaborators, we fail to nurture the very intellectual energy that enhances the scholarship in question. To take another example, when we only look to how scholarship is received and cited within the academy, we fail to reward and recognize the difficult, yet often more impactful work of participatory research oriented toward the goals and objectives of communities outside the academy itself.

Broad dashboards and frameworks that enable the quantitative assessment of scholars and scholarship might be useful to gain a general and superficial understanding of a wide scope of scholarly activity in relation to proxy indicators of quality. However, contextual, values-based frameworks—specific to institution, project, discipline, group, and individual—more effectively ensure that any assessment of the quality of scholarly practices and products is conducted using a rubric that aligns (a) with the core values or mission of the group and (b) with the values or professional motivations of the individual scholar. Ultimately, the values-aligned and values-enacted approach for which we advocate here at once more responsibly advances institutional missions and more effectively empowers individual scholars to undertake meaningful work of the highest quality that has a positive transformative impact on the lives of students, colleagues, and members of a wider public.

Developing a values framework

Until the onset of the coronavirus pandemic, our project team had been engaged in the provision of a series of day-long in-person workshops at research-intensive institutions in the United States designed to help academic faculty and staff—including

administrators, librarians, researchers, and others—create common, shared understandings of what they value as individuals and as institutions, whether that be, for example, engagement with a broader community, diversity and inclusion, open and transparent research processes, student-centered pedagogy, or experimental digital scholarship. At most institutions, a top-down values statement is often embedded, implicitly or explicitly, in college and university mission statements, but without any consideration of (a) what the socio-professional motivations and values of individual faculty members might be, (b) how values—institutional or personal—might be translated into the scholarly practice of individual members of the faculty, or (c) how institutions and disciplines recognize and reward scholarly work reflecting those values, often in very disproportionate ways. Our approach to working with these institutional teams in our workshops is threefold: (1) we facilitate discussions about the discrepancies between what a given institution claims to value and what they assess and reward; (2) we empower scholars and administrators to imagine what a values-aligned approach to assessment, reward, and recognition might look like at their institution; and then (3) we lead them through an exploration of how they might shift individual practice, departmental policies, school-level, and college-level decisions, and university procedures toward enacting and incentivizing those values. These interventions can take the form of personalizing tenure-and-promotion plans for individual members of the faculty, such as outlined in the discussion of the CPIL framework above, to creating values-aligned frameworks to guide future work and budget allocations.

Lessons learned from the October 2017 HuMetricsHSS Value of Values workshop (<https://humetricshss.org/blog/on-the-value-of-values-workshop-part-1/>) informed the attempt to integrate a values-enacted approach to institutional budget decision making in the College of Arts and Letters at Michigan State University. Drawing on the core values departments and programs had identified through facilitated values-conversations, the College identified equity, openness, and community as the core values that would shape the work of the College community (<https://cal.msu.edu/news/living-values/>). A commitment to putting these values into practice led to the development of an open and collaborative budget-request process in which all unit chairs and directors were asked to prepare two priority requests that were then shared with and evaluated by all the associate deans, chairs, and directors in the College. To facilitate the process, leaders across the College worked together to establish a budget request evaluation rubric that included consideration of values alignment, student success, faculty retention, and a variety of budgetary, resource, space, and time constraints. Having developed this rubric in an open and collaborative way, the associate deans, chairs, and directors evaluated one another's requests, and the results were mapped onto a matrix that provided a sense of the shared judgment of the College's leadership and informed the specific budgetary decisions made by the dean. Opening the budget request process to this collaborative effort at once deepened the understanding of the wide array of requests made across units and increased the number of collaborative requests received. Further, the shared articulation of the rubric the dean would use to make decisions enacted a commitment to transparency and deepened levels of trust among unit leaders and with the dean's office.

While these approaches to promoting values-enacted scholarly evaluation and institutional decision-making practices are promising, the HuMetricsHSS initiative is but one among a number of efforts seeking to align values more intentionally with outputs of scholarship. Other examples include the San Francisco Declaration on Research Assessment (DORA) (<https://sfedora.org/>), the UK's Responsible Metrics Initiative (<https://responsiblemetrics.org>), the

Leiden Manifesto (<http://www.leidenmanifesto.org/>), the European Network for Research Evaluation in the Social Sciences and the Humanities (ENRESSH) (<https://enressh.eu/>), EvalHum (<http://www.evalhum.eu/>), and the Research Evaluation Working Group within the International Network of Research Management Societies (INORMS) (<https://inorms.net/activities/research-evaluation-working-group/>), all of which have been working for years on the challenge of changing deeply entrenched academic evaluation cultures and promoting a more responsible approach to research evaluation processes.

DORA, a global effort spearheaded by scientists in 2013, has championed an approach to research evaluation that encourages the use of diverse research impact metrics. Notably, DORA has been signed by nearly 1400 organizations and more than 14,000 individuals who disavow the “use of journal-based metrics, such as Journal Impact Factors, in funding, appointment, and promotion considerations” and who recognize the “need to capitalize on the opportunities provided by online publication (such as ... exploring new indicators of significance and impact)”. DORA's prominence (owing to its association with respected research policy researchers, universities, and organizations like the Public Library of Science) has meant that the effort has loomed large in the minds of those engaging with all disciplinary evaluation practices. It was the first in a series of initiatives that aim to reform how research evaluation is practiced. Recently they have undertaken active efforts to build community engagement, including conducting webinar discussions and online forums and hosting monthly community calls.

The Responsible Metrics initiative was another such prominent effort, announced following the publication of the UK's Higher Education Funding Council for England's *Metric Tide* report (Wilsdon et al., 2015). The report explored the possible use of metrics in the 2014 Research Excellence Framework (REF2014) national evaluation exercise, and recommended that metrics should “support, not supplant” the role of expert peer review in evaluation for all disciplines, including the humanities and social sciences. According to the *Metric Tide* report, “responsible metrics can be understood in terms of a number of dimensions:

- Robustness: basing metrics on the best possible data in terms of accuracy and scope;
- Humility: recognizing that quantitative evaluation should support—but not supplant—qualitative, expert assessment;
- Transparency: keeping data collection and analytical processes open and transparent, so that those being evaluated can test and verify the results;
- Diversity: accounting for variation by field, and using a range of indicators to reflect and support a plurality of research and researcher career paths across the system;
- Reflexivity: recognizing and anticipating the systemic and potential effects of indicators, and updating them in response”. (pp. 134–135)

Each of these dimensions of Responsible Metrics is rooted in values similar to those shared by our project. The difference lies in that the concept of Responsible Metrics applies those values to the process of research evaluation itself, rather than to the values of the researcher per se.

Though the Responsible Metrics initiative is now largely inactive, a number of independent researchers and research administrators have taken up the “responsible metrics” mantle, publishing and speaking candidly on the challenges and opportunities inherent in the use of research indicators and implementing at their local universities the *Metric Tide* recommendations. In 2016–2018, for example, Durham University convened a Responsible Metrics Working Group charged

with addressing locally the recommendations of the Metric Tide report; their proposed framework was approved by the university senate on 16 October 2018 (<https://www.dur.ac.uk/library/research/evaluate/responsiblemetrics/>). In Summer 2019, Cardiff University advertised for a “responsible research metrics officer” (<https://thebibliomagician.wordpress.com/latestjobalerts/>), a position funded by the Wellcome Trust and eventually filled by Karen Desborough, Cardiff’s Responsible Research Assessment Officer (<https://www.cardiff.ac.uk/research/our-research-environment/integrity-and-ethics/responsible-research-assessment>). And others have written extensively about the concept (European Commission Expert Group on Altmetrics, 2017; Evans, 2018; McVeigh, 2018). These kinds of discussions and values-based approaches to research evaluation have the ability to shift the culture of higher education.

HSS-specific research evaluation projects are also influencing academia through their values-based practices, especially across Europe. ENRESSH is part of a pan-European intergovernmental framework that “aims to propose clear best practices in the field of SSH research evaluation” in a manner that reflects the diversity of HSS research. Along with the EvalHum consortium, which hosts an annual conference on HSS research evaluation, the work of ENRESSH to engage and serve researchers from all corners of Europe embodies the concepts of equity and diversity.

The INORMS Research Evaluation Working Group has developed SCOPE, a five-stage process for evaluating responsibly (<https://inorms.net/wp-content/uploads/2020/05/scope.pdf>): START with values, consider CONTEXT, weigh the OPTIONS for measuring, PROBE those options deeply, and EVALUATE the evaluation. The SCOPE model stresses that responsible assessment should be rooted in measuring what we as institutions value and care most about in research, while acknowledging that often “we either measure what others value, or measure using only the data that we have readily available”. The process of understanding and developing a contextualized value framework involves iterative steps intended to always probe and evaluate the evaluation process against the framework itself.

These initiatives all share an underlying desire both to recognize and encourage HSS research and pedagogy of the highest quality and to protect HSS scholars from the misuse and abuse of research impact indicators that only recognize a very narrow slice of scholarly labor in a system that is already highly focused on a very narrow slice of academic laborers and outputs. Building upon the efforts of other research assessment policy experts, however, we on the HuMetricsHSS initiative advance a unique “values-first” approach designed to transform the culture of higher education from within by holding individuals and institutions accountable to the values they identify as shaping their work.

Toward a taxonomy of values-enacted indicators

The current preference for evaluation metrics and the currency they use—citations—could be enriched simply by expanding the scope of sources for those indicators of quality and impact. This is already the work of altmetrics, defined by the National Information Standards Organization (2016) as “the collection of multiple digital indicators related to scholarly work...derived from activity and engagement among diverse stakeholders and scholarly outputs in the research ecosystem, including the public sphere”. Commercial altmetrics companies such as Plum Analytics and Almetric.com, as well as the nonprofit Our Research tool ImpactStory (<https://profiles.impactstory.org>), look for the digital trace of interest in research outputs in a wide variety of sources, including social media, library catalogs, clinical or policy documents, blog posts, and code repositories such as Github. At Plum Analytics (<https://plumanalytics.com/learn/about-metrics>), for

example, these “traces of interest” take five broad forms: citations, usage, captures (including bookmarks or forked code), mentions, and social media.

Where our approach differs from a standard altmetrics effort, however, lies in the intention and interpretation of our project. One of the aims of PlumX Metrics, for example, is to “enable analysis by comparing like with like”, while Almetric sells its services to researchers by promising that the data it provides “can be used to benchmark against other research published in your field, meaning you can see where the work of your peers is gaining traction” (<https://www.altmetric.com/audience/researchers/>). Such a focus on market analysis and peer competition can certainly be one measure of influence—possibly leading, as we have already argued, to toxic behavior in both individual and institutional contexts—but it does little to suggest how scholars are advancing work in alignment with their own individual and institutional goals and values. It also only supports, again, a summative approach to assessment.

A shift in the focus of an existing indicator or an expansion of its scope can, when adopted with intention, can be brought into alignment with a given values framework. For example, while citations in journals are (too) often (improperly) used as an indicator of the impact of a publication, citations within syllabi, as Cronin (2014) and Kousha and Thelwall (2014) argue, can demonstrate that a set of ideas is gaining in relevance and urgency for faculty and students.

Because the syllabus can function both as an artifact and an indicator of scholarship, it is an important locus of scholarly intervention and impact. As an artifact of scholarship, the syllabus offers faculty opportunities to embody the values about which they care most deeply through the texts they assign, the assignments they give, and the policies they adopt. We put it this way in a flash essay that argues that the syllabus is a locus of intervention and impact:

“As a construction site for notions of authority, legitimacy, and power, [the syllabus] proffers an opportunity to build up traditionally underrepresented voices and forms of scholarship and redefine the parameters of the scholarly conversation. As a form of sustained engagement with those voices, it offers a new way to conceive of scholarly impact that goes beyond just another citation”. (Agate et al., 2020)

If the construction of a syllabus invites scholars to put their values into practice through what and how they teach, the syllabus also serves as an indicator of scholarship for those scholars whose texts and scholarship are taught. Being cited in a syllabus indicates that the work is being engaged in a substantive way not only by a scholar’s peers, but by their students. This suggests that syllabus citations and the amount of time dedicated to a given artifact of scholarship could provide a richly textured account of the impact of a scholar’s work. As Jason Rhody (2016) puts it, “such indicators might help us rethink and influence notions of impact that currently favor an article-based intellectual economy”. (Inclusion of one’s work in a syllabus affords deep insight into impact: the level of the course, the institution in which it is taught, the amount of time dedicated to the scholarship, the assignments associated with the work—all become indicators of impact for scholars hoping to provide textured accounts of the extent to which their work is adopted and engaged.)

This focus on the syllabus as an indicator of scholarship moves from what is published to what is taught, from what scholars are reading to what students are learning, from the conference to the classroom. Here too our understanding of impact deepens, for when a given article or blog post or podcast is taught, we have access to a more textured understanding of its impact. We can discern where it is being taught, at which institutions and at what

levels; we can identify how much time is being dedicated to the ideas or if the content referenced is the focus of specific assignments and projects. Citations in syllabi also situate ideas within a broader network of scholarship that provides a deeper understanding of their impact. We might identify syllabus citations as an example of what we could call Expanded Scope and Deepened Focus Indicators.

A second set of examples of Expanded Scope and Deepened Focus Indicators can be discerned in an emerging new group of awards that are given in recognition of work that has for too long remained undervalued and unrecognized (Long, 2019). In 2019, for example, the journal *Language Testing* announced a new *Language Testing* Reviewer of the Year Award (<https://journals.sagepub.com/page/ltj/awards>). Established by co-editors Paula Winke and Luke Harding, the award recognizes a reviewer who has provided exceptional reviews according to the quality of the feedback provided. More specifically, the editors established criteria for the award associated with the values the journal wanted to cultivate. Reviews must provide knowledgeable feedback and open, resourceful advice; they must be collegial, kind, and timely; and the feedback must have a positive impact on the final publication. These criteria at once elevate the quality of the work published and recognize the importance of peer review and formative feedback in the creation of scholarship. Following this example, the *Journal for General Education* established its first Reviewer of the Year award in 2020 (http://www.psupress.org/Journals/jnls_jge.html).

A second type of values-enacted indicator might be called Vicarious Indicators. To foster a culture of generosity and—perhaps more importantly—to address the imbalance in the academic ecosystem that rewards only limited products (e.g., publications) that rely on substantial hidden labor, we can identify ways to recognize and reward faculty for facilitating or contributing to the success of their colleagues. For example, a mentor might point to very traditional indicators of success for a mentee—peer-reviewed journal publications, invitations to conferences, monograph publications, etc.—as indicators of effective mentoring. To avoid the very real danger of taking credit for the success of others, it is important to establish structures for mentees to certify and lend detail to the impact a mentor may have had on their mentee's success. The value of a culture of generosity, of course, is that it breaks the corrosive self-centered culture of competition that distorts relationships within and beyond the academy, and it also helps disrupt mythologies of individualism that pervade an academic culture that is replete with collegiality and collaboration, whether recognized or not. A “solitary scholar” is one who relies on an extensive network of colleagues in archives, libraries, and labs, and on deep reserves of peer feedback—through formative feedback mechanisms, from individual conversations and conference Q&A to formal peer review—that can often have transformative effects on their work, even if in small ways. While Vicarious Indicators might be too easily weaponized to take credit away from those who deserve it most, creating intentional structures and processes of reciprocal feedback can mitigate this danger even as it would enrich our understanding of the mutual process of creating most scholarly outputs.

Vicarious Indicators are already at work in the evaluation of academic administrators. Recognizing that effective administration requires a shift from the self-centered focus of the scholar to an other-oriented approach, the success of administrators is often measured by the success they empower others to achieve. These achievements are often discerned in traditional terms—research expenditures, fundraising success, faculty awards won, books or articles published, etc. Still, these indicators are vicarious because they depend on the success of others and are grounded in the extent to which that success can be traced to the conditions the

administrator helped establish to make them possible. Vicarious Indicators can also be deployed to capture the microtransactions of everyday collegiality and to recognize labor that is necessary but undervalued. As we discussed earlier, the increased scarcity of peer reviewers in the face of an overwhelming number of articles and books in need of review reflects a dangerously skewed reward system. But developing ways of capturing and rewarding contributions through formative and evaluative peer review would not only help rebalance this equation but might encourage more collegial, thoughtful, and encouraging feedback.

A third type of values-enacted indicator may be discerned in the way some academic units are beginning to integrate point systems into their annual review processes to give colleagues credit for activities that are valued by the group but not necessarily by the wider academic community. We might call these Values-Driven Quantification Indicators. The distinguishing feature of this type of indicator is that a group decides together to embrace a robust point system precisely to give weight to a variety of activities they deem worthwhile. The Department of Gender, Women's, and Sexuality Studies at the University of Iowa, for example, uses just such a system to count a broad range of contributions to teaching, research, and service the faculty have agreed to recognize as valuable not just to an individual's career but to the health of the department. The key to the successful adoption of such indicators is that the group has engaged in conversations that are intentionally structured to facilitate honest self-reflection about the values they share and how those values might themselves be valued in evaluating performance.

The transformative power of a values-enacted approach

If, as we have suggested, a first step in deepening the meaning of scholarly excellence is to realign our systems of evaluation with an explicit understanding of the values that shape our scholarship, we are now in a position to recognize that “quality” scholarship simply is intentionally enacting the values that give our work purpose. Notions of “excellence” when it comes to research evaluation have long been discussed and often contested (Adams and Gurney, 2014; Bernal and Villalpando, 2002; Brown and Leigh, 2018; Carli et al., 2019; Cremonini et al. 2018; Hamann, 2016; Hazelkorn, 2015; Hester, 2003; Hicks, 2012; Johnston, 2008; Kalpazidou Schmidt and Graversen, 2018; Kraemer-Mbula et al., 2020; Kwok, 2013; Ndofirepi, 2017; Oancea and Furlong, 2007; Thelwall and Delgado, 2015; Tijssen, 2003; Tijssen and Kraemer-Mbula, 2017; Vessuri et al., 2013). From Bill Readings' *The University in Ruins* (1997) and Michèle Lamont's *How Professors Think* (2010) to the more recent “Excellence R Us” article (Moore et al., 2016), we have repeated reminders that excellence often itself serves as a false proxy for evaluating scholarly work and institutions. Moore et al. suggest soundness as an alternative to excellence, but soundness has too limited a meaning associated with the structures of arguments to adequately capture the rich and textured expressions of quality found in a wide diversity of disciplines. Instead, our approach seeks to ground the meaning of excellence and quality in how well scholarship embodies a shared set of stated values appropriate to a given context and developed in genuine reciprocal dialog.

In short, we argue that to elevate the quality of the scholarship we produce, we must establish values-based frameworks that align with, recognize, and reward personal and institutional values enacted in supportive local contexts. Here the quality of scholarship depends not on the deployment of abstract standards that flatten the texture of our research, alienate us from the vitality of our work, and feed a corrosive culture of competition, but rather on empowering scholars to put their values into

intentional practice in meaningful ways. Shifting from abstract standards of excellence to enacting values-inflected purposeful work has the capacity to transform the culture of higher education—making scholarship more meaningful for those who undertake it, more valuable for the communities engaged by it, and more transparent to those who need to evaluate it.

Data availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during this study.

Received: 9 May 2020; Accepted: 28 October 2020;

Published online: 07 December 2020

Notes

- 1 Throughout this article we use the term “ecosystem” to refer to the interactions between scholars and the material environment in which they operate. Building, however, on Norris (2014), we would argue that when “ecosystem” is used to describe human-constructed social conditions then one needs also to interrogate the values that inform the construction of the systems themselves.
- 2 The toxic culture to which we refer here has a long history. Over a century ago, John Jay Chapman (1910) was already lamenting the way business practices had reshaped higher education since the 1870’s, with the result that professors had become more interested in protecting their own self-interest than in supporting one another in advancing the educational mission of the university.
- 3 See our discussion of this term and its critique on pages 13–14.
- 4 At the California State University at Fullerton, all students in a math class were required to purchase a textbook co-authored by the department chair and deputy chair (Jaschik, 2015). While that incident is a particularly egregious example of the practice, professors requiring students to purchase their textbooks is not uncommon, as Ian Ayres (2005) points out in arguing that college textbooks should be free to students, as they are in elementary and secondary school.
- 5 Inflation bias, also known as “p-hacking” or “selective reporting”, is the misreporting of true effect sizes in published studies. It occurs when researchers try out several statistical analyses or data eligibility specifications and then selectively report those that produce significant results.
- 6 It is important to note that while a values-based approach to academic labor is meant to recognize labor across the academy, including those not on the tenure track, the tenure clock and structures of tenured faculty advancement remain dominating constructs for university pacing, for those both on and off the tenure track.

References

- Aagaard K, Bloch C, Schneider JW (2015) Impacts of performance-based research funding systems: the case of the Norwegian publication indicator. *Res Eval* 24(2):106–117
- Adams J, Gurney KA (2014) Evidence for excellence: has the signal overtaken the substance? *Dig Sci*, London
- Agate N, Kennison R, Long CP, Rhody J, Sacchi S, Weber P (2020) Syllabus as locus of intervention and impact. *Syllabus* 9(1):300
- Ayres I (2005) Just what the professor ordered. *New York Times*. <https://www.nytimes.com/2005/09/16/opinion/just-what-the-professorordered.html>
- Baccini A, De Nicolao G, Petrovich E (2019) Citation gaming induced by bibliometric evaluation: a country-level comparative analysis. *PLoS ONE* 14(9): e0221212. <https://doi.org/10.1371/journal.pone.0221212>
- Baker KJ (2018) *Sexism Ed: essays on gender and labor in academia*. Bloomsbury, London
- Baker S (2018) How much research goes completely uncited? *Times Higher Education*. <https://www.timeshighereducation.com/news/how-much-research-goes-completely-uncited#survey-answer>
- Balietti S, Goldstone RL, Helbing D (2016) Peer review and competition in the Art Exhibition Game. *Proc. Natl Acad. Sci. USA* 113(30):8414–8419
- Bergeron D, Ostroff C, Schroeder T, Block C (2014) The dual effects of organizational citizenship behavior: relationships to research productivity and career outcomes in academe. *Hum. Perform.* 27(2):99–128
- Bernal DD, Villalpando O (2002) An apartheid of knowledge in academia: the struggle over the “legitimate” knowledge of faculty of color. *Equity Excell. Educ.* 35(2):169–180
- Bornmann L (2014) Do altmetrics point to the broader impact of research?: an overview of benefits and disadvantages of altmetrics. *J Informetr* 8(4):895–903
- Bouville M (2008) Plagiarism: words and ideas. *Sci. Eng. Ethics* 14(3):311–322
- Braxton JM, Proper EM, Bayer AE (2011) *Professors behaving badly: faculty misconduct in graduate education*. Johns Hopkins University Press, Baltimore
- Brown N, Leigh J (2018) Ableism in academia: where are the disabled and ill academics? *Disabil. Soc.* 33(6):985–989
- Carli G, Tagliaventi MR, Cutolo D (2019) One size does not fit all: the influence of individual and contextual factors on research excellence in academia. *Stud. Higher. Educ.* 44(11):1912–1930
- Chapman JJ (1910) *Professorial ethics*. *Science* 32(809):5–9
- Chubb J, Reed MS (2018) The politics of research impact: academic perceptions of the implications for research funding, motivation and quality. *Br.Pol.* 13(3):295–311
- Cilano C, Fritsche S, Hart-Davidson W, Long CP (2020) Staying with the trouble: designing a values-enacted academy. *Impact of Social Sciences*. <https://blogs.lse.ac.uk/impactofsocialsciences/2020/04/23/staying-with-the-trouble-designing-a-values-enacted-academy/>. Accessed 23 April 2020
- Clark A (2013) Haunted by images? Ethical moments and anxieties in visual research. *Methodol. Innov. Online* 8(2):68–81
- Cleary S (2018) APA member interview: Michaela McSweeney. *Blog of the American Philosophical Association*. <https://blog.apaonline.org/2018/04/20/apa-member-interview-michaela-mcsweeney>. Accessed 23 April 2020
- Collins PH (2012) Looking back, moving ahead: scholarship in service to social justice. *Gen. Soc.* 26(1):14–22
- Cremonini L, Hurlings E, Hessels LK (2018) Different recipes for the same dish: comparing policies for scientific excellence across different countries. *Sci.Pub. Policy* 45(2):232–245
- Cromey DW (2010) Avoiding twisted pixels: ethical guidelines for the appropriate use and manipulation of scientific digital images. *Sci. Eng. Ethics* 16(4):639–667
- Cronin B (2014) Scholars and scripts, spoors and scores. In: Cronin B, Sugimoto CR (eds) *Beyond bibliometrics: harnessing multidimensional indicators of scholarly impact*. MIT Press, Cambridge, p 3–21
- Davenport E, Snyder H (1995) Who cites whom? whom do women cite?: an exploration of gender and scholarly citation in sociology. *J Doc* 51(4):404–410
- Dean KL, Forray JM (2018) The long goodbye: can academic citizenship sustain academic scholarship? *J Manag. Inq.* 27(2):164–168
- DeChavez Y (2018) It’s time to decolonize that syllabus. *Los Angeles Times*. <https://www.latimes.com/books/la-et-jc-decolonize-syllabus-20181008-story.htm>
- Diamantes T (2005) Online survey research of faculty attitudes toward promotion and tenure. *Essays Educ.* 12(1):3, <https://openriver.winona.edu/eie/vol12/iss1/3> <https://openriver.winona.edu/eie/vol12/iss1/3>
- Díaz-Faes AA, Bordons M, van Leeuwen TN (2016) Integrating metrics to measure research performance in social sciences and humanities: the case of the Spanish CSIC. *Res. Eval.* 25(4):451–460
- Douglas RJ (1992) How to write a highly cited article without even trying. *Psych. Bull.* 112(3):405–408
- European Commission Expert Group on Altmetrics (2017) *Next-generation metrics: responsible metrics and evaluation for open science*. European Commission Directorate-General for Research and Innovation, Brussels
- Evans K (2018) What difference does a responsible metrics statement make? *The Bibliomagician*. <https://thebibliomagician.wordpress.com/2018/10/24/what-difference-does-a-responsible-metrics-statement-make/>, Accessed 23 April 2020
- Fernandez-Llmos F (2019) Peer review and publication delay. *Pharm. Practice* 17(1):1502
- Fitzpatrick K (2019) *Generous thinking: a radical approach to saving the university*. Johns Hopkins University Press, Baltimore
- Flaherty C (2016) Refusing to be measured. *Inside Higher Education*. <https://www.insidehighered.com/news/2016/05/11/rutgers-graduate-school-faculty-takes-stand-against-academic-analytics>
- Franck G (1999) Scientific communication—A vanity fair? *Science* 286(5437):55
- Gelman A, Loken E (2013) The Garden of Forking Paths: Why Multiple Comparisons Can Be a Problem, Even When There Is No ‘Fishing Expedition’ or ‘p-Hacking’ and the Research Hypothesis Was Posited Ahead of Time. 14 November, http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf. Accessed 20 March 2020
- Gingras Y (2016) *Bibliometrics and research evaluation: uses and abuses*. MIT Press, Cambridge
- Grossberg M (2004) Plagiarism and professional ethics: a journal editor’s view. *J Am* 90(4):1333–1140
- Gruber T (2014) Academic sell-out: how an obsession with metrics and rankings is damaging academia. *J Market. Higher Educ* 24(2):165–177
- Hamann J (2016) The visible hand of research performance assessment. *Higher Educ* 72(6):761–779
- Hammarfelt B, de Rijcke S (2015) Accountability in context: effects of research evaluation systems on publication practices, disciplinary norms, and

- individual working routines in the faculty of arts at Uppsala University. *Res. Eval.* 24(1):63–77
- Hansson SO (2008) Philosophical plagiarism. *Theoria* 74(2):97–101
- Hatherhill JA (2018) *Questioning bibliometrics and research impact*. [conference presentation] Bibliometrics and Research Impact Conference, Ottawa, Canada. <http://hdl.handle.net/10393/38100>
- Hazelkorn E (2015) Rankings and the reshaping of higher education: the battle for world-class excellence, 2nd ed. Palgrave Macmillan, London
- Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD (2015) The extent and consequences of *p*-hacking in science. *PLoS Biol* 13(3):e1002106. <https://doi.org/10.1371/journal.pbio.1002106>
- Held V (2015) Care and justice, still. In: Engster D, Hamington M (Eds.) *Care ethics and political theory*. Oxford University Press, Oxford, p 19–36
- Hester TL (2003) Choctaw conceptions of the excellence of the self, with implications for education. In: Waters A (Ed.) *American Indian thought: philosophical essays*. Wiley–Blackwell, Malden, p 182–187
- Hicks D, Wouters P, Waltman L, de Rijcke S, Rafols I (2015) Bibliometrics: the Leiden Manifesto for research metrics. *Nature* 520(7548):429–431
- Hicks D (2012) Performance-based university research funding systems. *Res. Policy* 41(2):251–261
- Hojat M, Gonnella JS, Caellegh AS (2003) Impartial judgment by the “gatekeepers” of science: fallibility and accountability in the peer review process. *Adv. Health Sci. Educ* 8(1):75–96
- Hüppauf B (2018) A witch hunt or a quest for justice: an insider’s perspective on disgraced academic Avital Ronell. *Salon*. <https://www.salon.com/2018/09/08/a-witch-hunt-or-a-quest-for-justice-an-insiders-perspective-on-disgraced-academic-avital-ronell/>
- Jaschik S (2015) Can a professor be forced to assign a \$180 textbook? PBS Newshour. <https://www.pbs.org/newshour/education/can-professorforced-assign-180-textbook>, Accessed 8 Oct 2020
- Johnston R (2008) On structuring subjective judgements: originality, significance and rigour in RAE2008. *Higher Educ. Quart.* 62(1):120–147
- Jordan S (2014) Research integrity, image manipulation, and anonymizing photographs in visual social science research. *Int Journal Soc Res Methodol* 17(4):441–454
- Kaiser Z (2016) Our program. *Mediated Space*. <http://mediated.space/Our-Program>. Accessed 23 April 2020
- Kalpazidou Schmidt E, Gravarsen EK (2018) Persistent factors facilitating excellence in research environments. *Higher Educ* 75(2):341–363
- Kousha K, Thelwall M (2014) Web impact metrics for research assessment. In: Cronin B, Sugimoto CR (Eds.) *Beyond bibliometrics: harnessing multidimensional indicators of scholarly impact*. MIT Press, Cambridge, p 289–306
- Kraemer-Mbula E, Tijssen RJW, Wallace ML, McLean R (eds.) (2020) *Transforming research excellence: new ideas from the global south*. African Minds, Cape Town
- Kulkarni S (2016) Manipulating the peer review process: why it happens and how it might be prevented. *Impact of Social Sciences*. <https://blogs.lse.ac.uk/impactofsocialsciences/2016/12/13/manipulating-the-peer-review-process-why-it-happens-and-how-it-might-be-prevented/>. Accessed 23 April 2020
- Kwok JT (2013) Impact of ERA research assessment on university behaviour and their staff. National Tertiary Education Union, Melbourne
- Lamont M (2010) *How professors think: inside the curious world of academic judgment*. Harvard University Press, Cambridge
- Laudel G, Gläser J (2006) Tensions between evaluations and communication practices. *J Higher Educ Policy Manag* 28(3):289–295
- Lawrence PA (2002) Rank injustice. *Nature* 415(6874):835–836
- Lee CJ, Schunn CD (2011) Social biases and solutions for procedural objectivity. *Hypatia* 26(2):352–373
- Long CP (2019) Recognizing peer review. *HuMetricsHSS*. <http://humetricshss.org/blog/recognizing-peer-review>
- Macfarlane B (2007) Defining and rewarding academic citizenship: the implications for university promotions policy. *J Higher Educ Policy Manag* 29(3):261–273
- Mahoney MJ (1977) Publication prejudices: an experimental study of confirmatory bias in the peer review system. *Cogn Ther* 1(2):161–175
- Mallard G, Lamont M, Guetzkow J (2009) Fairness as appropriateness: negotiating epistemological differences in peer review. *Sci Tech Hum* 34(5):573–606
- Martin B (1997) Academic credit where it’s due. *Campus Review* 7(21):11
- McKiernan EC, Schimanski LA, Nieves C, Matthias L, Niles MT, Alperin JP (2019) Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations. *PeerJ Preprints* 7:e27638v2, <https://peerj.com/preprints/27638>
- McVeigh M (2018) Metaphor and metrics. *Web of Science Group*. <https://clarivate.com/webofsciencegroup/article/metaphor-and-metrics/>. Accessed 23 April 2020
- Moher D, Naudet F, Cristea IA, Miedema F, Ioannidis JPA, Goodman SN (2018) Assessing scientists for hiring, promotion, and tenure. *PLOS Biol* 16(3):e2004089. <https://doi.org/10.1371/journal.pbio.2004089>
- Moore S, Neylon C, Eve MP, O’Donnell DP, Pattinson D (2016) “Excellence R Us”: university research and the fetishisation of excellence. *Pal Commun* 3:105. <https://doi.org/10.1057/palcomms.2016.105>
- Mounk Y, Bergstrom CT, Smith JEH, Petrzel NM, Schieber D, Heying HE, Essig L, Moorti S (2018) A *Chronicle* forum: What the “grievance studies” hoax means. *The Chronicle of Higher Education*. <https://www.chronicle.com/article/What-the-Grievance/244753>
- Mountz A, Bonds A, Mansfield B, Loyd J, Hyndman J, Walton-Roberts M, Basu R, Whitson R, Hawkins R, Hamilton T, Curran W (2015) For slow scholarship: a feminist politics of resistance through collective action in the neoliberal university. *ACME* 14(4):1241, <https://www.acmejournal.org/index.php/acme/article/view/1058> <https://www.acmejournal.org/index.php/acme/article/view/1058>
- National Information Standards Organization Altmetrics Initiative Working Group (2016) *Outputs of the NISO Alternative Assessment Metrics Project*. National Information Standards Organization, Baltimore, Report number RP-25-2016
- Ndofirepi A (2017) African universities on a global ranking scale: legitimization of knowledge hierarchies? *South African J Higher Educ* 31(1):155–174
- Niles MT, Schimanski LA, McKiernan EC, Alperin JP (2020) Why we publish where we do: faculty publishing values and their relationship to review, promotion and tenure expectations. *PLoS ONE* 15(3):e0228914. <https://doi.org/10.1371/journal.pone.0228914>
- Norris T (2014) Morality in information ecosystems. *CLIR Connect*. <https://connect.clir.org/blogs/tim-norris/2014/09/18/morality-in-informationecosystems>. Accessed 7 Oct 2020
- Noy S, Ray R (2012) Graduate students’ perceptions of their advisors: is there systematic disadvantage in mentorship? *J Higher Educ* 83(6):876–914
- Oancea A, Furlong J (2007) Expressions of excellence and the assessment of applied and practice-based research. *Res. Paper Educ.* 22(2):119–137
- Onbekend N, Waltman L, de Rijcke S (2016) What do we know about journal citation cartels?: a call for information, CWTS Meaningful Metrics. <https://www.cwts.nl/blog?article=n-q2w2b4>. Accessed 22 April 2020
- Penders B (2018) Ten simple rules for responsible referencing. *PLoS Comput. Biol.* 14(4):e1006036. <https://doi.org/10.1371/journal.pcbi.1006036>
- Pontille D, Torny D (2010) ‘The controversial policies of journal ratings: evaluating social sciences and humanities’. *Res. Eval.* 19(5):347–360
- Rafols I, Leydesdorff L, O’Hare A, Nightingale P, Stirling A (2012) How journal rankings can suppress interdisciplinary research: a comparison between innovation studies and business and management. *Res. Policy.* 41(7):1262–1282
- Readings B (1997) *The University in Ruins*. Harvard University Press, Cambridge
- Resnik DB (2012) Plagiarism: words and ideas. *Account Res* 19(5):269–272
- Rhody J (2016) The syllabus as HuMetrics case study. *HuMetricsHSS*. <https://humetricshss.org/blog/the-syllabus-as-humetrics-case-study/>. Accessed 7 Oct 2020
- Rouhi S (2017) When metrics and politics collide: reflections on peer review, the JIF and our current political moment. *The Scholarly Kitchen*. <https://scholarlykitchen.sspnet.org/2017/09/18/guest-post-metrics-politics-collide-reflections-peer-review-jif-current-political-moment/>. Accessed 27 March 2020
- Sabarathnam M, Kirby P. (2014) Why metrics cannot measure research quality: a response to the HEFCE consultation. *The Disorder of Things*. <https://thedisorderofthings.com/2014/06/16/why-metrics-cannot-measure-research-quality-a-response-to-the-hefce-consultation/>. Accessed 21 March 2020
- Schliesser E (2018) What do hoax papers really show?: ten theses. *Digressions and Impressions*. <https://digressionsimpressions.typepad.com/digressionsimpressions/2018/10/what-do-hoax-papers-really-show-h.html>. Accessed 22 April 2020
- Scott N (2012) Altmetrics are the central way of measuring communication in the digital age but what do they miss? *Impact of Social Sciences*. <https://blogs.lse.ac.uk/impactofsocialsciences/2012/12/17/scott-altmetrics-central-digital-whats-missing/>. Accessed 22 April 2020
- Solnit R (2014) *Men explain things to me*. Haymarket Books, Chicago
- Sugimoto CR (2014) Academic genealogy. In: Cronin B, Sugimoto CR (Eds.) *Beyond bibliometrics: harnessing multidimensional indicators of scholarly impact*. MIT Press, Cambridge, p 365–382
- Thelwall M, Delgado MM (2015) Arts and humanities research evaluation: no metrics please, just data. *J Document* 71(4):817–833
- Tijssen RJW (2003) Scoreboards of research excellence. *Res. Eval.* 12(2):91–103
- Tijssen RJW, Kraemer-Mbula E (2017) Perspectives on research excellence in the global south: assessment, monitoring and evaluation in developing country contexts. Science Granting Councils Initiative, Nairobi
- Vessuri H, Guédon J-C, Cetto AM (2013) Excellence or quality? Impact of the current competition regime on science and scientific publishing in Latin America and its implications for development. *Curr Sociol* 62(5):648–665
- White HD (2004) Reward, persuasion, and the Sokal Hoax: a study in citation identities. *Scientometrics* 60(1):93–120

- Wicherts JM, Veldkamp CLS, Augusteijn HEM, Bakker M, van Aert RCM, van Assen MALM (2016) Degrees of freedom in planning, running, analyzing, and reporting psychological studies: a checklist to avoid *p*-hacking. *Front Psychol* 7:01832. <https://doi.org/10.3389/fpsyg.2016.01832/full>
- Wilsdon J, Allen L, Belfiore E, Campbell P, Curry S, Hill S, Jones R, Kain R, Kerridge S, Thelwall M, Tinkler J, Viney J, Wouters P (2015) *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. Higher Education Funding Council for England, Bristol, Report number RG.2.1.4929.1363
- Zacchia Z (2017) The dark side of discrimination in the economics profession: how women are forced to conform to the research habits and interests of men. <https://www.ineteconomics.org/perspectives/blog/the-dark-side-of-discrimination-in-the-economics-profession>. Accessed 23 April 2020

Acknowledgements

Many thanks to HuMetricsHSS team members Bonnie Thornton-Dill and Bonnie Russell for their contributions to this work.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to C.P.L.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020