Viewpoint | David Lorge Parnas

Stop the Numbers Game
Counting papers slows the rate of scientific progress.

As a senior researcher, I am saddened to see funding agencies, department heads, deans, and promotion committees encouraging younger researchers to do shallow research. As a reader of what should be serious scientific journals, I am annoyed to see the computer science literature being polluted by more and more papers of less and less scientific value. As one who has often served as an editor or referee, I am offended by discussions that imply that the journal is there to serve the authors rather than the readers. Other readers of scientific journals should be similarly outraged and demand change.

The cause of all of these manifestations is the widespread policy of measuring researchers by the number of papers they publish, rather than by the correctness, importance, real novelty, or relevance of their contributions. The widespread practice of counting publications without reading and judging them is fundamentally flawed for a number of reasons:

It encourages superficial research. Those who publish many hastily written, shallow (and often incorrect) papers will rank higher than those who invest years of careful work studying important problems; that is, counting measures quantity rather than quality or value;

It encourages overly large groups. Academics with large groups, who often spend little time with each student but put their name on all of their students' papers, will rank above those who work intensively with a few students;

It encourages repetition. Researchers who apply the "copy, paste, disguise" paradigm to publish the same ideas in many conferences and journals will score higher than those who write only when they have new ideas or results to report;

It encourages small, insignificant studies. Those who publish "empirical studies" based on brief observations of three or four students will rank higher than those who conduct long-term, carefully controlled experiments; and

It rewards publication of half-baked ideas. Researchers who describe languages and systems but do not actually build and use them will rank higher than those who implement and experiment.

Paper-count-based ranking schemes are often defended as "objective." They are also less time-consuming and less expensive than procedures that involve careful reading. Unfortunately, an objective measure of contribution is frequently contribution-independent.

Proponents of count-based evaluation argue that only good papers get into the "best" journals, and there is no need to read them again. Anyone with

Viewpoint

experience as an editor knows there is tremendous variation in the seriousness, objectivity, and care with which referees perform their task. They often contradict one another or make errors themselves. Many editors don't bother to investigate and resolve; they simply compute an average score and pass the reviews to the author. Papers rejected by one conference or journal are often accepted (unchanged) by another. Papers that were initially rejected have been known to win prizes later, and some accepted papers turn out to be wrong. Even careful referees and editors review only one paper at a time and may not know that an author has published many papers, under different titles and abstracts, based on the same work. Trusting such a process is folly.

Measuring productivity by counting the number of published papers slows scientific progress; to

Anything goes. Researchers publish things they know may be wrong, old, or irrelevant; they know that as long as the paper gets past some set of referees, it counts;

Bespoke research. Researchers monitor conference and special-issue announcements and "custom tailor" papers (usually from "pre-cut" parts) to fit the call-for-papers;

Minimum publishable increment (MPI). After completing a substantial study, many researchers divide the results to produce as many publishable papers as possible. Each one contains just enough new information to justify publication but may repeat the overall motivation and background. After all the MPIs are published, the authors can publish the original work as a "major review." Science would advance more quickly with just one publica-

If you get a letter of recommendation that counts numbers of publications, rather than commenting substantively on a candidate's contributions, ignore it.

increase their score, researchers must avoid tackling the tough problems and problems that will require years of dedicated work and instead work on easier ones.

Evaluation by counting the number of published papers corrupts our scientists; they learn to "play the game by the rules." Knowing that only the count matters, they use the following tactics:

Publishing pacts. "I'll add your name to mine if you put mine on yours." This is highly effective when four to six researchers play as a team. On occasion, I have met "authors" who never read a paper they purportedly wrote;

Clique building. Researchers form small groups that use special jargon to discuss a narrow topic that is just broad enough to support a conference series and a journal. They then publish papers "from the clique for the clique." Formation of these cliques is bad for scientific progress because it leads to poor communication and duplication, even while boosting the apparent productivity of clique members;

tion; and

Organizing workshops and conferences. Initiating specialized workshops and conferences creates a venue where the organizer's papers are almost certain to be published; the proceedings are often published later as a book with a "foreward" giving the organizer a total of three more publications: conference paper, book chapter, and foreward.

One sees the result of these games when attending conferences. People come to talk, not to listen. Presentations are often made to nearly empty halls. Some never attend at all.

Some evaluators try to ameliorate the obvious faults in a publication-counting system by also counting citations. Here too, the failure to read is fatal. Some citations are negative. Others are included only to show that the topic is of interest to someone else or to prove that the author knows the literature. Sometimes authors cite papers they have not studied; we occasionally see irrelevant citations to papers with titles that sound relevant but are not. One can observe researchers improving both their

publication count and citation count with a sequence of papers, each new one correcting an error in the hastily written one that preceded it. Finally, the importance of some papers is not recognized for many years. A low citation count may indicate a paper that is so innovative it was not initially understood.

Accurate researcher evaluation requires that several qualified evaluators read the paper, digest it, and prepare a summary that explains how the author's work fits some greater picture. The summaries must then be discussed carefully by those who did the evaluations, as well as with the researcher being evaluated. This takes time (external evaluators may have to be compensated for that time), but the investment is essential for an accurate evaluation.

A recent article [1], which clearly described the methods used by many universities and funding agencies to evaluate researchers, offered software to support these methods. Such support will only make things worse. Automated counting makes it even more likely that the tactics I've described here will go undetected.

One fundamental counting problem raised in [1] is the allocation of credit for multiple-author papers. This is difficult because of the many author-ordering rules in use, including:

Group leaders are listed first, whether or not they contributed;
Group leaders are listed last, whether or not they contributed.
Authors are listed in order of contribution, greatest contribution first;
Authors are listed by "arrival," that is, the one who wrote the first draft is first; and
Authors are listed alphabetically.

Attributing appropriate credit to individual authors requires either asking them (and believing their answers) or comparing the paper with previous papers by the authors. A paper occasionally contributes so much that several authors deserve full credit. No mechanical solution to this problem can be trusted. It was suggested in [1] that attention be restricted to a set of "leading" journals primarily distinguished by their broad coverage. However, there are often more substantive and important contributions in specialized journals and conferences. Even "secondary" journals publish papers that trigger an important new line of inquiry or contribute data that leads to a major result.

Only if experts read each paper carefully can they determine how an author's papers have contributed to their field. This is especially true in computer science where new terms frequently replace similar concepts with new names. The title of a paper may make old ideas sound original. Paper counting cannot reveal these cases.

Sadly, the present evaluation system is self-perpetuating. Those who are highly rated by the system are frequently asked to rate each other and others; they are unlikely to want to change a system that gave them their status. Administrators often act as if only numbers count, a probability because their own evaluators do the same.

Those who want to see computer science progress and contribute to the society that pays for it must object to rating-by-counting schemes every time they see one being applied. If you get a letter of recommendation that counts numbers of publications, rather than commenting substantively on a candidate's contributions, ignore it; it states only what anyone can see. When serving on recruiting, promotion, or grant-award committees, read the candidate's papers and evaluate the contents carefully. Insist that others do the same. **C**

REFERENCE
1. Ren, J. and Taylor, R. Automatic and versatile publications ranking for research institutions and scholars. *Commun. ACM 50,* 6 (June 2007), 81–85.

DAVID LORGE PARNAS is Professor of Software Engineering and Director of the Software Quality Research Laboratory in the Department of Computer Science and Information Systems at the University of Limerick, Limerick, Ireland.