



Dangers of Rankings with Inaccurate Data

By Eric Grimson, CRA Board Chair, on behalf of the CRA board

Our culture is embedded with rankings: of movies, of college athletic teams, of consumer products, of universities, and of graduate programs. Rankings are a guilty pleasure—we claim they don't influence us, and we know their foibles, yet we can't help looking to see where we stand.

Academics understand the problems behind reputational rankings such as the US News and World Report's ranking of universities, of graduate and undergraduate programs, and of specialties within a field: they are largely subjective and influenced by non-scientific factors, they have long time-constants and are subject to hysteresis, and they at best reflect an overall assessment of a program without acknowledging exceptional elements. Yet we also know that rankings are used by prospective students, by university administrators allocating limited resources between units, and by sponsors. When the NRC announced that it would update its 1995 ranking of graduate programs, many hoped this would provide a complementary perspective on doctoral programs as the charge to the NRC (portions excerpted in italics below¹) stressed a quantitative basis:

An assessment of the quality and characteristics of research-doctorate programs in the United States will be conducted. The study will consist of: 1) the collection of quantitative data through questionnaires administered to institutions, programs, faculty, and admitted to candidacy students (in selected fields); 2) the collection of program data on publications, citations, and dissertation keywords; and 3) the design and construction of program ratings using the collected data including quantitatively based estimates of program quality. These data will be released through a web-based, periodically updatable database and accompanied by an analytic summary report.

In addition to releasing the data and ranges of rankings based on statistical sampling of weights for different measures of impact, the NRC would also provide tools to allow personalized ratings by adapting weights.

In principle, this is an excellent idea: it extracts detailed quantitative data (e.g., the current career track of every doctoral graduate from the previous five years, and publications from curricula vitae of faculty members), it provides tools to determine ranges of ratings rather than a single number, and it articulates specific measures for estimating impact (publications per faculty member, average citations per publication, percent of faculty with grants, and awards per faculty member).

Such a rating system, however, critically depends on its data acquisition process. If publications and citations are accurately estimated, this system may provide a quantitative alternative to

reputational ratings. Thus, the CRA board was very disappointed to hear at our February meeting that the NRC methodology for counting publications and citations appears to be badly flawed. The NRC chose to use the following methodology²:

Method of collecting publications, citations, and awards – *With the exception of fields in the humanities, publications and citations were collected through the Institute for Scientific Information (ISI), now a part of Thomson Scientific, and matched to faculty lists for fields in the sciences (including the social sciences). Although faculty were also asked about their publications in Section D of the faculty questionnaire, these lists were used only to check the completeness of the ISI data. The citation count is for the years 2000-2006 and relates to papers published between 1981 and 2006.*

As CS researchers know well, our field relies heavily on refereed conference publications, rather than the journal publications common in more traditional scientific domains. Indeed, tenure and promotion cases often depend heavily on conference publications (such as SIGGRAPH in graphics, or STOC or FOCS in theory, or ICCV or CVPR in computer vision). Unfortunately, the data acquired by the NRC through ISI (now known as Thomson Reuters) misses many of the publications we value:

- According to the NRC, only conference publications that also appeared as issues of a journal were included in the Thomson Reuters data. While Thomson Reuters now includes some top conferences in its listings, during the time period of 2000-2006, it did not. Thus virtually all conference publications and associated citations are missing from the NRC study.
- In computer science, publication venues evolve quickly; yet Thomson Reuters may take years before even a journal is included in its data. A listing of sources from NRC's data acquisition includes only 12 of the 34 transactions and journals published by the ACM in the 2000-2006 period; even now, Thomson Reuters includes only 4 of 8 ACM journals and 22 of 31 transactions. Thus, journal publications may also be poorly represented in the NRC data.

Even when the NRC did collect publication data, its categorizations are puzzling. For example, researchers may be surprised to learn that Computer Vision and Image Understanding is categorized under Language and Linguistics, and *International Journal of Computer Vision* is categorized under Aerospace Engineering; or that *ACM Transactions on Database Systems* is categorized under AI, Robotics and Automatic Control, as are Communications of the ACM and *ACM Transactions on Programming Languages and Systems*. While publications in these venues were included in the data acquisition, one wonders about the accuracy of such a system.

When the NRC solicited input as it began planning for the new ratings system, both the CRA and the ACM responded—each group independently urged the NRC to recognize that computer science researchers have shifted to publications in conferences over journals. In fact, the CRA provided a copy of its White Paper on assessing publications for promotion and tenure cases to the NRC in support of this argument. The NRC’s own 1994 study, *Academic Careers for Experimental Computer Scientists and Engineers*, noted: “A substantial majority of respondents to the CRA-CSTB survey of ECSE faculty preferred conferences as the means of dissemination by which to achieve maximum intellectual impact; many fewer preferred journals.” The input from the CRA (on behalf of its member societies) and from the ACM was provided in 2002, yet 8 years later it appears that the NRC still has not grasped the importance of conference publications in our field. And if they are not measuring conference publications, they are thus also missing huge numbers of citations, both to conference articles and in conference articles to other papers.

The CRA board is thus concerned that the data being used to assess research impact of computer science programs are flawed. If one is simply trying to compare programs within computer science, then under-sampled data might not be an issue, since all programs are subject to the same data acquisition methods. However, this assumes one has a representative sampling of publications and citations; since subfields have different publishing practices, this is simply not valid. Moreover, when one compares assessments of computer science programs against other fields, which we believe many university administrators will do, the underestimate of research impact implicit in this flawed data may badly misrepresent CS programs, and support poor decisions on resource allocations that affect those programs.

As of the writing of this article, the NRC has not yet released its ratings system, and they may well be wrestling with this quandary. In fact, the CRA board sent a detailed letter to the NRC (and to the NAE, NAS, IOM, and the chairman of the committee responsible for the program assessment) indicating our concerns with these data and with the damage that their release may have on the field. We have not yet heard from the NRC on its plans, so in anticipation of that release, we suggest some possible remedies.

One option is to release the NRC ranking system and data with a disclaimer indicating the problems with the citation and publication data, and a discussion of how NRC will remedy this in future releases. While this would allow the NRC to release a report that is already many years late in completion, the CRA board urges the NRC not to proceed with this approach. We are concerned that once the data are released, they will simply be used without any acknowledgement of the disclaimer, and will thus support invalid conclusions about the relevance of computer science as a discipline.

If the NRC does not release the flawed data, what are other options? An alternative would be not to release any ranking of computer science programs until the NRC can complete a

scientifically accurate assessment of computer science research productivity and impact. In this approach, there would be no premature release of partial data that could be misinterpreted, and there would be an incentive for NRC to complete the correct acquisition and analysis of data in order to complete their report. We are concerned, however, that releasing a ranking system for doctoral programs that does not include computer science may be damaging to our field. While many institutions view computer science as an important element of science and engineering research and education, there are still institutions in which computer science struggles for acceptance and respect. Hence, the CRA board is concerned that not including computer science in the ranking system may damage the field's standing at some institutions, leading to reduced resource allocation and other effects. Given the recent U.S. Bureau of Labor Statistics report (<http://www.bls.gov/>) showing that computer science graduates earn higher- than-average salaries, that employment growth in computer science is expected to be much faster than average, and given their projection that computing occupations are likely to grow by 22.2 percent between now and 2018 (the fastest growing cluster of all professional occupations), the CRA board fears that undercutting computer science departments would not only damage the field, but could have significant impact on the nation's economy.

If not releasing the computer science part of the report is problematic, then **an alternative is to release the computer science ranking report without any citation data (and including a clear statement from the NRC explaining why the data are absent and how the NRC will remedy this in future releases), but including complete publication data.** To ensure that complete, accurate publication data are included, we strongly urge the NRC to work with the CRA and its member societies. We believe we can provide to the NRC an assessment of relevant journals and of conferences that our member societies consider to be strongly refereed.

Thus, the CRA is prepared to help the NRC define an accurate list of publishing venues, from which they could then determine publishing statistics for computer science faculty members from the curricula vitae submitted in the NRC data acquisition process. This option, presuming that publication data are deemed to provide a reasonable, though partial, assessment of research productivity, would allow the release of the ranking system since it covers three of the four metrics proposed by the NRC for faculty assessment (publications per faculty member, average citations per publication, percent of faculty with grants, and awards per faculty member). If the NRC cannot obtain accurate publication data, we would urge them to delay publication until they can remedy this situation.

Even if accurate publication data can be obtained, the issue of measuring citations is still unresolved. The CRA strongly encourages the NRC to work with Thomson Reuters or some other partner, such as Google Scholar, to implement a citation assessment system that fully measures the computer science field. This means capturing refereed conference

publications and associated citations, and not focusing solely on journal articles. Once these data are acquired, we urge the NRC to reissue their ratings system and report.

In summary, the CRA board is very concerned about releasing flawed citation and publication data as part of the NRC report. The board has urged the NRC to acquire accurate publication data and to release a ranking system that includes those data but does not include the flawed citation data; if that is not possible, the board urges the NRC to withhold the release of computer science programs until they can fairly assess the productivity and impact of those programs.

Finally, the board urges the NRC to work with CRA, Thomson Reuters, and other partners to implement an accurate and fair citation measurement system for computer science. Without this, we are concerned that the NRC's ratings system will incorrectly portray the field of computer science in a manner damaging to all of us.

Eric Grimson, Chair of CRA's Board of Directors, is a Professor of Computer Science and Engineering at the Massachusetts Institute of Technology, and holds the Bernard Gordon Chair of Medical Engineering at MIT. He also holds a joint appointment as a Lecturer on Radiology at Harvard Medical School and at Brigham and Women's Hospital.

End Notes:

1 From: http://sites.nationalacademies.org/PGA/Resdoc/PGA_044479

2 From: "A Guide to the Methodology of the National Research Council Assessment of Doctoral Programs," available at: http://www.nap.edu/catalog.php?record_id=12676

Editorial Note:

The CRA Board provided copies of a version of this article to the Presidents of the NAE, NAS, and IOM, to the chairman of the committee responsible for the graduate program assessment, and to the study director. The CRA also distributed copies to many of the computer scientists who are members of the NAE or NAS, and gathered reactions from those members.

As this newsletter was going to press, the CRA Board was contacted by the NAE. After extended discussions between the CRA and the NAE, and internally within the NRC, the CRA Board received a formal response from the NRC. In that response, the NRC acknowledges the validity of the concerns raised in the article. The NRC states that it will work with the CRA to acquire more accurate publication data, including journal and conference publications, by extracting data from submitted CVs of faculty members; it will not include any citation data in its released report; it will delay the report release until the publication data can be acquired and assessed; and it will include in its report an acknowledgement of the missing citation data and an articulation of its plan for acquiring that data for future releases.

Finally, the NRC promises to work with the CRA and its member societies to develop an improved system to gather publication and citation data. The CRA appreciates the NRC's willingness to acknowledge the problems with the initial system, and its willingness to work with the CRA and its member societies to create a more valid system for rating doctoral programs.