National Academy Panel Finds 'Health' of Science Fields Defies Measurement

Accountability is among the latest watchwords in Washington. But a group of scientists at the National Research Council who attempted to figure out some sort of accountability for Federal funding within and across scientific fields found the task all but impossible.

Accountability was a centerpiece in the Government Performance and Results Act of 1993 (PL 103-62), which requires all Federal agencies to establish their respective goals and to measure their progress toward achieving those goals. In passing the legislation, Congress expected that the resulting metrics would be used to evaluate the performance of the departments and agencies as a way of examining their health and to serve as a guide to make them more accountable. In a totally unrelated effort, House Speaker Newt Gingrich's "Contract with America" calls for the legislators to abide by the same civil rights and employment laws that apply to the rest of the country. Accordingly, the first piece of legislation passed by the House in the first 24 hours of the 104th Congress was HR 1, a bill to make members of Congress accountable for the same hiring and employment practices that apply to everyone else. After the bill was approved by the House, the Senate passed a similar measure, and President Clinton signed it into law (PL 104–1) on 23 January.

The issue of accountability in science has a much different context. For the research council, which operates under the aegis of the National Academies of Sciences and Engineering and the Institute of Medicine, accountability was rarely—if ever—a subject for the sweeping surveys that examined the status and directions of scientific fields and subfields in the past. Such studies include "Physics Through the 1990s," a survey of the entire field issued in 1986 and called the Brinkman Report (after the committee's chairman, William F. Brinkman of AT&T Bell Labs), "Opportunities in Chemistry," the 1985 review known as the Pimentel Report (after the late George C. Pimentel of the University of California, Berkeley), and "The Decade of Discovery in Astronomy and Astrophysics," released in 1991 and often referred to as the Bahcall Report (for John Bahcall of the Institute for Advanced Study in Princeton, New Jersey). All three reports identified promising research and led in some instances, particularly in astronomy, to increases in Federal funding. Even so, the science community hasn't found ways of setting priorities across different fields or subfields or of assessing the vitality of a field or its accountability in spending Federal money to sustain its "health." Indeed, when a special planning committee organized by the American Physical Society about seven years ago attempted to do that, the panel couldn't agree on what standards and values to apply.

The latest effort, conducted by the research council's Commission on Physical Sciences, Mathematics and Applications, began about 18 months ago with the belief that data could be collected on three fields—namely, mathematics, astronomy and astrophysics, and atomic, molecular and optical sciences. The commission assumed at the outset that the data would be similar for all fields and thus enable it to measure accountability. According to the panel's chairman, Richard N. Zare, a physical chemist at Stanford University, "We decided that if we could assess the health of various fields, then it might be possible to compare fields to see which were in better shape and which needed help. So we started looking for various indicators of health, like the success of a field in attracting students or the number of times publications in a certain field are cited."

Areas of potential metrics

Besides collecting data on such matters, Zare's group identified several other broad areas of potential metrics: the capability of a field to attract funds for graduate programs, research activities and equipment or facilities; the quantitative output of high-quality students and professionals (with emphasis on women and underrepresented minority groups); the field's adaptability in adjusting to changes in scientific opportunities, levels of support and national needs; the effectiveness of the field in setting priorities and optimizing the use of resources; and the contribution of the field to world science and to society in general.

It soon became clear to the panel that the Federal R&D agencies report their data differently, depending upon their respective missions and their particular way of labeling programs for political advantage. Worse, "each discipline has its own standards or vital signs for measuring health," says the panel's report, "Quantitative Assessments of the Physical and Mathematical Sciences," issued on 27 December. So the metrics "must be relative and not absolute, [and] even within a single discipline the metrics and their interpretation will vary depending on who is doing the assessment and for what purpose."

As Zare writes in the preface to the report: "It is easy to posit that what is bigger is better, but when the bigger must become smaller, the affected community often will marshal measures to show that the health of the field is deteriorating when the contrary may actually be the case. . . . It is seductive to imagine that a set of metrics exist that, when tracked over time, will enable predicting with confidence the future prospects of a scientific field or subfield. Whereas some indicators may be useful for judging the normal pace of progress in a field as it follows its established paradigm, no evidence could be found that suggests such statistical measures have long-range predictive power. How can we know whether some supposedly dull field in stasis or even in decline is on the verge of a remarkable rebirth that will result from an unexpected experimental discovery or from some new and vital idea coming from the work of just one individual?

"We must not let a cycle be created in which the need for accountability leads to the use of measurement standards, which leads to polishing existing paradigms, which leads to further demands for accountability, and so on. . . . Put another way, we must avoid 'looking for the lost key where the light is brightest': We must avoid allowing what can be measured to become what matters, rather than seeking to measure what matters, which frequently are attributes that cannot be measured."

Zare does not argue against accountability. "Anyone who uses public funds has to be accountable," he says. Statistical measures of performance need to be combined with expert judgments, he observes. And while the claims of experts should be treated with skepticism, he notes, "the experts can themselves be held accountable, and to ignore their expertise is to ignore the best performance assessments anyone can devise." So the panel's experiment to find metrics that Congress and others, including scientists themselves, might use to assess the accountability of a field was a failure. Nonetheless, says Zare, "the findings should be reported and discussed as both a warning and a guidepost for future efforts to judge the effectiveness of public investments in science."

▶IRWIN GOODWIN ■