Scientometrics Vol. 7. Nos 1–2 (1985) 29–38

A FRAMEWORK FOR ASSESSING AND REWARDING A SCIENTIST'S RESEARCH PRODUCTIVITY

R. A. LEARY

School of Forestry and Wood Products Department of Forestry, Michigan Technological University, Houghton, Michigan 49931 (USA)

(Received March 27, 1984)

A contest for world leadership in science and technology exists. New ways to motivate scientists seem as important to contest outcome as new sources of funds. A framework formed by cross-tabulating question difficulty and answer generality should help to identify the contribution of a research scientist. A reward relationship based on this framework should help to ensure that scientists will work on the most difficult research problems, a necessity for a high quality research program.

Introduction

It is generally recognized that in the coming decades there will be severe competition for world leadership in science and technology. A commonly held belief is that federal governments can ensure a competitive, if not leading, position for their country by infusing large sums of money into science education, and science and technology research. Additional funds for science education can ensure the potential for leadership, but more funds for research will not ensure a leading position. The contest is not likely to be decided solely at the level of science administration and funding. Ignored in these long-range plans are the people who must actually do the work, make the discoveries, take the risks–the scientists. Ultimately, the contest for scientific and technological leadership will be decided in the laboratories, fields, and minds of scientists. Developing new ways to motivate scientists seems just as important as new sources of funds. Crucial to the motivation of scientists is the manner in which their productivity is assessed and the way their assessment is translated into reward.

American university and federal research scientists are evaluated on both a yearly and periodic basis. Periodic evaluations typically involve a scientist's peers. In universities peer evaluations are normally conducted twice in a career: for promotion from assistant to associate to full professorships. Periodic peer evaluations in federal research laboratories are supposed to occur no less frequently than every three or four years, or 6 to 8 times in a career. They are conducted according to the research grade evaluation guide of the Office of Personnel Management. Written in 1964 and revised slightly in 1967, the research grade evaluation guide is used in evaluating tens of thousands of American federal research scientists. Three outcomes of these panel evaluations are possible: demotion, retention in grade, or promotion.

Yearly evaluations are normally done for allocation of annual wage increments, and are conducted by administrators; department heads at universities and research work unit leaders in federal agencies. Little is written about criteria used by university administrators, although it is reasonable to assume that some measure of research productivity is a part of their evaluation equation. Federal yearly evaluations are conducted according to principles spelled out in the merit pay provisions of the Civil Service Reform Act of 1978[1]. In carrying out the provisions of the act, great emphasis is being given to research productivity defined as the number of publications per unit time. Publications quotas are being set for federal scientist's critical performance elements, thereby laying the ground for possible dismissal should the promised publications not be forthcoming by the set dates[2]. Higher quotas are given to scientists with higher GM grades.

Several questions exist in the way a legitimate concern for enhanced productivity and accountability is being carried out. First, is research productivity a scientific question or an administrative one? So-called modern methods of research management often require no knowledge on the part of administrators of the scientific disciplines being handled. Second, is scientific productivity being 'measured' with an acceptable unit of measure? Number of publications seems misguided[3]. Third, did the fine art of fragmenting research result into units just large enough to warrant publication, so-called least publishable units (LPU)[4] result from an over-emphasis on counting? Fourth, what are the likely consequences of fragmenting research results in least publishable units given the already difficult task of ensuring that research results are applied[5]? Fifth, has the practice of giving scientists publication quotas forced scientists to work on small, easy problems, the solution of which will contribute little to the contest over world scientific leadership? In sum, will an emphasis on measuring its scientist's research productivity by counting number of publications have a negative effect on scientific research in the United States, both in federal agencies and universities? I believe it will, and suggest a simple framework that will allow a scientist's research output to be evaluated without emphasis on the number of publications in which the accomplishments are communicated.

Parts of the research process

Research is a many-faceted enterprise. *Bunge*[6] identifies and contrasts two basic classes of research problems: substantive or object research problems and procedure (method) or strategy problems. Procedural or methodological problems are extremely important. Many scientists of necessity engage in techniques research. Early scientists were foreced to construct apparatus, glassware, and analytical devices to conduct valid experiments. Even today, for some scientists methodology is an end in itself. But for most it is a means to an end–making statements about reality. I limit my discussion here to the evaluation of object research.

Object scientific research can be thought to have three parts:

1. A scientifically stated question about some aspect of reality.
2. The methods and techniques by which the question is resolved.
3. A statement of the answer to the question.

Because I do not consider the problem of evaluating methodologies, the methods and techniques by which the question is answered are, if valid, not of concern in assessing productivity. I adopt the Feyerabendian notion that anything goes![7] What counts for purposes of measuring productivity in solving object research problems are the question and the answer. In support of this decision, a scientist with a truly superior method or technique will be able to make more meaningful statements about reality than a scientist without the technique. Thus, superior methods and techniques developed by a scientist will be manifest in more profound statements about reality. This leads to a first axiom of research productivity assessment:

Axiom 1: A unit of scientific productivity is a valid answer to a scientific question.

Thus, the fundamental unit of productivity is an answer to a single scientific question. A scientific innovation will typically involve answering several interrelated questions. But innovations should not be measured by counting the number of questions answered in their development. This would simply be counting on another level.

The question

In a given research area within a scientific discipline at a given time, there will likely be a number of questions of varying difficulty to answer. For example, questions of biomass accumulations in forests may be framed in a number of manners. "What is the amount of standing crop biomass in northern hardwood forests in the Lake States? " is a question that, conceptually at least, can be answered quite simply with sample surveys. If the question had been, "What amount of standing crop biomass could there be if all northern hardwood forests in the Lake States were fully stocked? ", it would be of greater difficulty to answer. One must know the relation between forest stocking and biomass accumulation as well as the state of stocking in existing forests. Another level of difficulty is "Why does biomass accumulation in northern hardwood forests never exceed about 325 metric tons per hectare?"[8] To answer this question one must integrate knowledge for several scientific disciplines. This leads to a second axiom of scientific productivity assesment:

Axiom 2: Not all scientific questions are of the same difficulty to answer.

"What is the amount?", or in general "What is?" questions can be answered with verbal descriptions and numerical tabulations. "What if?" questions are answered with predictions and retrodictions. "Why?" questions are answered with explanations. A third axiom of productivity assessment:

Axiom 3: With respect to a particular problem, "Why? " questions are more difficult to answer than "What if?" questions, which are more difficult to answer than "What?" of "What is?" questions.

Astute observation reported in the form of description has been, and will continue to be, an important source of new research opportunities. Narrative description, along with tabulations of objective numerical data, is a point of departure for science. Highly successful scientists are often keen observers who have the intellect to progress beyond description to likely causes. "Research is to see what everybody has seen and think what nobody has thought".[9]

"What if?" questions are answered by appealing to a family of cognitive functions based on anticipation and retrospection. *Bunge*[10] includes expectation, guessing, prophecy, prognosis, and scientific prediction (and retrodiction) in the family. Scientific prediction can be separated into purely statistical ("best fit") predictions and that based on laws of nature.

Answers to "Why?" questions, explanations, are complex items[11]. Explanations have two parts: explanandum (the fact being explained), and the explanans (that which explains). It is the structure and content of the explanans that is sought when answering "Why?" questions. The structure may, in the simplest cases consist of a generalization and a single circumstance as

Not every discipline is at a stage of development where "Why?" questions can be answered. The social sciences are often suggested to be in this condition, and medical science may also fit this state of development[12]. Nor is every scientist at a stage in his/her career to tackle the most difficult kind of question that is answered with an explanation. This leads to a fourth and fifth axiom of scientific research productivity assessment:

Axiom 4: Scientific disciplines are at different stages of development as given by the class of question typifying research in the discipline.

Axiom 5: Scientists within disciplines are at different stages of personal development as evidence by the class of question they answer successfully.

The answer

Just as research questions come in all sorts of forms, answers may be in different forms. It is, however, convenient to think of the answer to a scientific question as the meaning of a single declarative sentence – a proposition. *Bunge*[6] states that a proposition is a definite statement that can be true or false. Statement definiteness is ensured by complete specification of object and predicate variables in the answering proposition. Truth or falsity of a proposition can be ensured by affixing to the proposition a quantifier. Thus, answering propositions may be true in one case (singular), or prefixed with existential quantifiers such as "in at least one case" (indefinite existential), "in n cases" (definite existential), "in all cases in universe A" (bounded universal), or "in every case" (unbounded universal). Generally applicable answering propositions are often given law level stature in a hypothesis–law–theory spectrum of scientific constructs[13]. Suggested are two more axioms of research productivity assessment:

Axiom 6: Answering propositions come with varying kinds of existential quantifiers or universes.

Axiom 7: Answering propositions with more general quantifiers (universal or bounded universal) have a better chance of becoming law level statements then singular answering propositions or ones with existential quantifiers (indefinite or definite).

Assessment framework

A simple framework for assessing the research productivity of a scientist, in a particular discipline at a particular time, working on a particular class of problems, can be constructed by cross-tabulating question and answer forms (Fig. 1). Across the top are listed general question forms in order of increasing difficulty to answer. Down the left margin are listed answering propositions in order of increasing generality. Answers of the type associated with rows near the top come from singular, and indefinite and definite existentially quantified propositions. Near the bottom are bounded universal answering propositions. Answers of the latter type may be bounded in a number of ways, e.g., taxonomically, geographically, and ecologically. The bottom row is for answering propositions with a universal quantifier, law level statements.

Two major productivity gradients are evident in the body of the framework. Descriptions (left-most column) contain a gradient from those expressed as singular propositions to those quantified indefinitely to universal propositions (line A, Fig. 1).

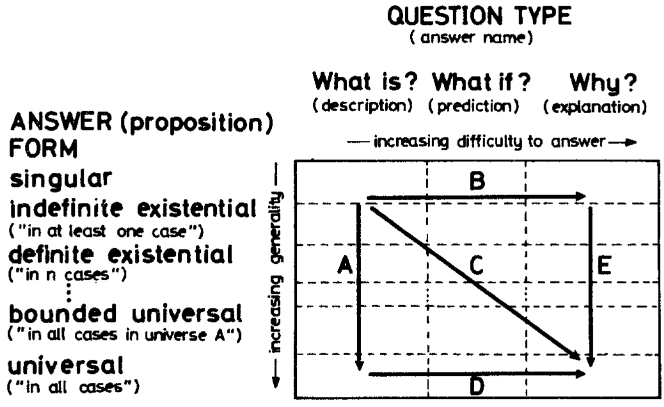

Fig. 1. Framework for evaluating the research productivity of a scientist by cross-tabulating question difficulty and answer generality. Meanings of arrows A–E are discussed in the text

Generally applicable descriptions represent greater productivity than isolated descriptions. Similarly, explanations of great generality represent greater productivity than those of little generality (line E).

Answers to "Why? " questions represent greater productivity than descriptions or predictions/retrodictions of the same level of generality (lines B and D, Fig. 1). Normally, if one has an explanation, it is possible to provide a prediction/retrodiction or a description. Not so the converse. Of course, there are gradations within categories such as predictions from polynomial based models with no clear interpretation for variable combinations or numerical constants to process-based mathematical predictions. The latter may have more the flavor of an explanation than the former, even though both are in the category I call prediction.

Because there are two gradients in the table, difficulty to answer across the table and generality of the answer down the table, it follows that a coarse grid is formed that identifies cells that can be given a productivity unit.

A grid for assessing productivity

Two problems must be addressed. First, productivity units must be assigned to cells formed by the grid, i.e., the grid must be calibrated. Calibration may be only in a qualitative or ordinal sense, but for purposes of discussion I suggest a quantitative calibration. Second, uses of the grid by research institutions, supervisors, and leaders must be identified and developed.

Calibration

Most important in the assignment of productivity units are the gradients from top to bottom rows, and from left-most to right-most columns. Bottom row cells should be assigned higher productivity units than top row cells. But how much higher? Twice? Four times? If assigned only twice the value, scientists may determine that they can accumulate more productivity units by developing answers to several special cases. Clearly, row D cells must have productivity units several times those of row B.

Column E cells should be given more productivity units than column A cells because explanations are more valuable than descriptions. They are more rare, which means they are more difficult to develop. Thus, column E cells should be given some multiple of the productivity units in column A. But what multiple?

Calibration of the complete grid may be accomplished by mathematical composition of sets of marginal values:

p = difficulty ⊗ generality,

where p = designates research productivity.

If, for example, one takes the composition function (⊗) to be ordinary multiplication, and the gradient from left-most column to right-most is 10, and from top row to bottom row is 10, a universal explanation constitutes 100 times the productivity of a singular description[14]. The steepness of the gradients is, again, a scientific question, not an administrative one. In the next section I look at whether a scientist will be rewarded in a way than preserves this hypothetical differential.

Uses of a performance grid

Science is a social enterprize. The reward a scientist receives for research that provides an answer occurring in a particular cell of Fig. 1 is very much a case of institutional style. Suggested is another axiom:

Axiom 8: Reward $\propto p^s$

where s designates an institutional style exponent (0,1)

p designates the productivity unit (difficulty ⊗ generality)

$\propto$ denotes 'is directly proportional to'.

The exponent for 'counting' research institutions will be near zero. Every unit is given the same weight. 'Excellence' research institutions will have an exponent at least as large as one, perhaps larger.

Some important practical uses of the grid and reward formula follow. First, the formula in axiom 8 signals the kind of research an institution seeks and will reward. Institutions (and administrators in institutions) with s values near zero will no doubt get mostly singular or indefinite descriptions. They are not only easier, they are less risky. Institutions with s values near, or greater than, one will probably get attempts at bounded or universal explanations. Second, the large number of 'publications' claimed by some scientists often makes it difficult for others to evaluate true ability as a researcher. The vitae of skimmers, Abstract and Proceedings artists, and LPUers invariably look and feel better than those of deep thinkers. They are thicker and heavier. But what do they contain in productivity units? Third, truly innovative research administrators would expect more productivity units in a set time period from higher paid scientists than from those receiving less, not more publications per year. If a universal explanation can be communicated in a single paper or monograph, why force a scientist to dilute the literature with pieces to make his meal? Fourth, the proliferation of scientific literature is not a proliferation of new knowledge. *Ney*[15] observed that in 1960 one year's issues of the *Astrophysical Journal* occupied 3 inches of shelf space in his office. In 1983 the Journal issues required two feet of shelf space. *Ney* argues that if the growth rate continues, by the year 200C one year's issues of the Journal will occupy 16 feet of shelf space. A spelling out of values for s and p by research institutions and administrators might dampen the perceived necessity to publish prolifically. Fifth, some institutions are administratively organized so that universal, or even bounded universal, answering propositions cannot be achieved. For example, forest growth research in the U. S. Forest Service research branch is administratively divided into about 15 research work units spread across 4 federal forest experiment stations in the eastern United States. The forests of concern to each research group are administratively protected from other research groups, thereby virtually ensuring that no law-level-like propositions will be developed that express the growth processes of all eastern forests. Sixth, the scientific productivity in an innovation is close to the sum of the productivity scores of answers comprising the innovation. Seventh, professional development of a research scientist could be qualitatively charted on the grid. Some young scientists may begin with indefintely quantified descriptions. Fine. But whole careers should not be spent on them[16]. Some scientists may attempt to develop explanations to account for a few instances of a phenomenon, hoping to generalize the finding with additional research. They would take the B to E route to universal explanations. Others might attempt to develop descriptions of great generality hoping to frame them in such a way that predictions and explanations can be developed. They would take the A to D route to universal explanations. Still other scientists may gamble by going directly for universal explanations. Perhaps the normal course of development of a scientific discipline is akin to arrow C in Fig. 1. Cautious advice to a young scientist may be to pattern a career after it.

In sum, science will be better served if the emphasis on measuring research productivity by counting a scientist's publications is rapidly swept aside as an administrative fad. Highly motivated scientists need administrators who go beyond mere numerosity in their performance evaluations. The grid formed by question difficulty and answer generality is a first step in this direction. Research institutions can help by letting existing as well as prospective scientists know their reward style. Doing so may provide a better match of personal capabilities and goals with institutional willingness to reward risk and excellence. Unless changes are made in how research productivity is assessed, a nation's scientists may each be highly 'productive', yet that nation may not compete successfully for world scientific and technological leadership.

References and Notes

1. United States Statutes at Large, 1978. Volume 92, Public Law 95–454, Statute 1111.
2. I searched both the House of Representatives and Senate publications containing testimony on the Civil Service Reform Act and found no reference or comment about the effect of the Act's merit pay provisions on federal research scientists.
3. Even the widely known studies of *Pelz* and *Andrews* assessed a scientist's productivity by counting the number of publications, patents or patent applications, and unpublished reports in a five year period. Neither question difficulty nor answer generality played a role in their assessment procedure. D. PELZ F. ANDREWS, *Scientists in Organizations,* John Wiley, New York, 1966.
4. W. BROAD, *Science* 211 (4487) (1981) 1137–1139.
5. General Accounting Office. The Forest Service needs to ensure that the best possible use is made of its research findings, Report Number B–125053, 1972.
6. M. BUNGE, *Scientific Research I: The Search for System,* Springer-Verlag, New York, 1967.
7. P. FEYERABEND, *Against Method–Outline of an Anarchistic Theory of Knowledge, NLB, London, 1975.*
8. *G. MROZ, private communication.*
9. A. SZENT–GYÖRGYI, *Bioenergetics,* Academic Press, New York, 1957.
10. M. BUNGE, *Scientific Research II: The Search for Truth,* Springer-Verlag, New York, 1967.
11. C. HEMPEL, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science,* Free Press, New York, 1965; E. KLEMKE, R. HOLLINGER, A. KLINE, Eds, *Introductory Readings in the Philosophy of Science,* Prometheus Books, Buffalo, 1980.
12. G. MYRDAL, *Bulletin of the Atomic Scientists,* 29 (1) (1973) 31–37; L. THOMAS, *The Youngest Science: Notes of a Medicine Watcher,* The Viking Press, New York, 1983.
13. J. HOSPERS, *An Introduction to Philosophical Analysis,* Prentice-Hall, Englewood Cliffs, 1967. I limit myself to answers as single statements. Thus, I do not include theories, which may consist of several law-like statements[8]. Generality of applicability or domain of truth is but one of several criteria that can be used to select from among several candidate law statements. Others that have been suggested are: formal simplicity, approximation to the truth, and theoretical tractability.

14. H. NAHIKIAN, *A Modern Algebra for Biologists,* The University of Chicago Press, Chicago, 1964.
15. E. NEY, *Science* 222 (4623): 456.
16. A. SZENT-GYÖRGYI, *Perspectives in Biology and Medicine* 15 (1) (1971) 1–5. *Szent-Györgyi* even argues that older scientists should tackle the more difficult problems, thereby leaving some of the less difficult to younger scientists in need of some successes at the start of their career.
17. I thank E. *Bakuzis,* R. *McRoberts,* E. H. T. *Whitten,* G. *Brown,* R. *Lee,* and R. *Monserud* for constructive review comments.