27 February 1981, Volume 211, Number 4485 SCIENCE

Innovation and Evaluation

Frederick Mosteller

Summary. Social, medical, and technological innovations are discussed, first with reference to historical examples and then with modern studies. I show the need for evaluating both the innovations themselves and the research processes leading to them. I suggest some kinds of research that need to be carried out if we are to continue to have a vigorous program of scientific and technological innovation. Finally, I explain the new initiative by the AAAS in science and engineering education.

My topic here is innovation and evaluation. I begin with an early experiment in nutrition. It was designed by Daniel of the Lions' Den, but for humans rather than lions. Daniel was held hostage in Nebuchadnezzar's court and, possibly for religious reasons, disapproved of the rich food, wine, and meat served there. The eunuch in charge feared for his own head if he were to give Daniel and his three friends merely the simple Judean vegetable fare called pulse (such as peas and beans). Daniel asked for a 10-day trial and promised to turn to the court's diet if the Judean hostages weren't then as healthy as the others. To turn to a translation of the original article, Daniel I:12-15 (*1*):

> Prove thy servants, I beseech thee, ten days; and let them give us pulse to eat, and water to drink.
>
> Then let our countenances be looked upon before thee, and the countenance of the children that eat of the portion of the king's meat: and as thou seest, deal with thy servants.
>
> So he [the eunuch] consented to them in this matter, and proved them ten days.
>
> And at the end of ten days, their countenances appeared fairer and fatter in flesh than all the children which did eat the portion of the king's meat.

Had this study been submitted as a report to *Science*, the reviewer might make the following remarks. First, there is no sampling problem because Daniel needed only to prove that he and his three friends were better off with the diet. He did not have to make the generalization to, say, the entire population of Judea or the human race. This is unusual because ordinarily we *are* trying to make such generalizations. For Daniel it was fortunate as well, because with such a small sample—Daniel, Shadrach, Meshach, and Abednego—the eunuch would have had to insist on using Student's t-test, and this would not be invented for another 2500 years, almost exactly.

Second, the length of the trial, 10 days, seems short for a nutrition experiment.

Third, the end point "fairer and fatter in flesh" seems not well defined. Other translations speak of "sleeker" which also is vague.

From the eunuch's point of view, the diet of pulse was an innovation, while the court's regular diet was the standard. And so Daniel designed a comparative experiment, an early evaluation of an innovation.

I turn to a historical, but more policy-oriented example: Another nutrition experiment was carried out by James Lancaster starting in 1601 when the East India Company sent its first expedition to India. He was general of four ships and a victualler (*2*). They sailed from Torbay in England in April 1601. At that time scurvy was the greatest killer of the navy and of expeditions and explorations, worse than accidents or warfare or all other causes of death together. More than half a crew might die of scurvy on a long voyage. In 1497 Vasco da Gama sailed around the Cape of Good Hope with a crew of 160 men: 100 died of scurvy (*3*).

Lancaster served three teaspoons of lemon juice every day to the sailors on the largest ship of his fleet and few became ill. By the time the fleet got to the Cape of Good Hope, so many sailors on the three smaller ships were sick from scurvy that Lancaster had to send sailors from the large ship to rig the smaller ones. When they reached the Cape of Good Hope 110 men had died, mostly from the 278 men who started on the three smaller ships. Clear evidence that lemon juice prevents scurvy? Maybe. At any rate, the evidence is so strong that the East India Company and the British Navy could surely be expected to follow up this investigation with further research. Not at all! Policy moves more majestically.

About 150 years later, 1747 to be precise, the physician James Lind (*4*) carried out an experiment consisting of adding something special to the diets of scurvy patients on the ship *Salisbury*. He had six dietary additions:

1) Six spoonfuls of vinegar.
2) Half-pint of sea water.
3) Quart of cider.
4) Seventy-five drops of vitriol elixir.
5) Two oranges and one lemon.
6) Nutmeg.

Lind assigned two sailors ill from scurvy to each treatment. Those who got the citrus fruit were cured in a few days and were able to help nurse the other patients. The supply of citrus fruit ran out in about 6 days.

Lind knew about Lancaster's work as well. With this dramatic and crucial experiment plus the backup of Lancaster's earlier voyage surely the British Navy

Frederick Mosteller is Roger I. Lee Professor, Department of Biostatistics, School of Public Health, Harvard University, Boston, Massachusetts 02115. He has just retired as president of AAAS. This article is the text of his address at the annual meeting of the AAAS in Toronto, Ontario, Canada, 6 January 1981.

0036-8075/81/0227-0881$01.50/0

now will adopt citrus fruit for prevention of scurvy from long sea voyages? No! Forty-eight years later policy caught up. In 1795 the British Navy began using citrus juice on a regular basis and wiped out scurvy in the service (*5*). The British Board of Trade followed suit after a delay of only 70 years (1865) and wiped it out in the mercantile marine (*5*). We often talk about how slow we are to make use of innovations, but this case study of citrus juice should give us a little encouragement. Today we are worrying about 20-year lags. Here is one of 264 years.

Evaluation of Today's Social Programs

We need both to have a larger number of innovations and to be sure that they are beneficial. This requires both inventiveness and evaluation. Let me demonstrate first with social programs.

To see whether social programs that had been evaluated carefully were successful, Gilbert, Light, and I (*6*) reviewed a substantial number of programs. Each had been evaluated by a randomized controlled experiment. I mention only a few, to give you a feeling for their variety.

After we studied the evaluation of a program, we scored the program on a scale running from a double plus down to a double minus, with zero meaning there was essentially no gain from the program, a double plus meaning that the program was an excellent innovation, and a double minus that it was much worse than the treatment that it replaced. Our ratings did not include costs of the program; had they done so, they would probably have had to be reduced somewhat.

The studies were classified into social, sociomedical, and medical innovations. Four of the eight social innovations we studied were: negative income tax, studies of bail, training of police, and attempts to reduce delinquency among girls.

Let me describe one. The study of delinquent girls was intended to reduce the amount of juvenile delinquency by instituting a social program. It had two steps. First, the investigators needed to identify the potentially delinquent girls, and second to apply the program to them and so prevent their delinquency.

What happened in the experiment? First, the innovators were very successful in identifying those girls who were *likely to become* delinquents. Second, they had no success at all in diverting the young girls from their course. Thus we assigned this innovation a zero. Although it is worth something to be able to identify potential delinquents, and this feature would be useful in future studies, the purpose of the innovation was to reduce delinquency. Since it did not do this, it was rated a zero. It did not increase the delinquency rate either, and so it did not get either a minus or a double minus.

We also studied eight sociomedical innovations, of which four were: experiments on probation for conviction for public drunkenness, effects of broadening health insurance benefits, training mothers whose children have tonsillectomies, and training physicians in comprehensive medical care.

Let me briefly describe the probation experiment. The judge assigned these habitual offenders to one of three treatments in a randomized manner. Such offenders with two arrests in the previous 3 months or three in the previous year were fined $25, given a 30-day suspended sentence, and assigned to one of three groups:

1) No treatment.
2) An alcoholic clinic.
3) Alcoholics Anonymous.

The payoff variables were number of rearrests and time before first rearrest. The results were that the "No treatment" group performed somewhat better than the other two groups, each of which performed equally well.

The original authors, Ditman *et al.* (*7*), concluded that the study gave no support to the policy of short-term referrals. Thus we scored the innovation a zero. It might possibly have been scored a minus.

In this same group was the Kansas-Blue Cross experiment. It had been suggested that one reason for excessive use of hospitalization and consequently of the rising costs of medical care was that the insurers would pay only for work done in the hospital, whereas some work could be taken care of more cheaply with outpatient care. The insurance company responded to this suggestion with a substantial randomized experiment. They put 5000 people into a group that had added benefits of ambulatory care, free of charge, in addition to the regular hospitalization, and compared their results over a year with those of 10,000 patients on the regular program.

The results came out contrary to expectations. The amount of hospitalization for the group with extra ambulatory benefits went up by 16 percent, while that for the group with regular benefits increased by only 3 percent. Thus the overall effect went in the opposite direction to that hoped for. There was more information. For example, there were 15 percent fewer short-term stays in the extra-benefit group. But the decrease was more than offset by the added longer stays. This innovation received a minus because the results went in the direction opposite to that hoped for.

It is, of course, possible that there may be a benefit overall in this approach because of finding things wrong early that need attention and thus preventing later health problems. That would take a substantial further investigation to establish.

We studied 12 medical innovations, including the following four: the Salk vaccine experiment for polio prevention, treatment of cancer of the bronchus, vagotomy, and gastric freezing for ulcer.

The Salk vaccine was a major success and has nearly stamped out paralytic polio. In one investigation, children were allocated to two groups, those injected with the vaccine and those injected with a saline solution. The vaccine was highly effective, and we rated it a double plus.

The treatment for cancer of the bronchus presented a difficult rating problem because on the one hand there was a substantial improvement in survival, but on the other we did not have clear evidence about the quality of the lengthened life. The comparison was between surgery and radiation therapy. Patients receiving the radiotherapy survived almost 50 percent longer—up from about 200 days to nearly 300, so we gave the treatment a double plus. But a plus might have been appropriate because further information about quality of life might have changed our views.

These few examples illustrate some experiments and reforms and their evaluation.

How did it all come out?

Out of 28 innovations, 12 were positive, and of these 6 got double pluses, 3 were negative, and 13 were rated zero. Thus less than half of well-tested innovations we discovered in the literature were beneficial even if costs were neglected. This suggests strongly that social or medical innovations *do* need to be evaluated.

You may be concerned that our sample of innovations was rather catch-as-catch-can. We were troubled about this too, and therefore did a further study.

We evaluated surgical innovations (*8*). By using the MEDLARS search system, a computer-based bibliography of medical literature, we obtained a population of surgical studies that was selected in an objective way.

We chose randomized clinical trials. In all, we found for the period under

study—1964 to 1973—a total of 36 trials comparing a surgical innovation against a standard treatment. Among these, 44 percent were regarded as successful. Of these, 11 percent were not improvements over the standard, but they were the equal of it and offered new approaches that might be preferred in special circumstances. The actual improvements were about 33 percent, of which the excellent ones comprised 14 percent.

Thus we find again that when innovations are put to trial, they are successes only about half the time, and that substantial improvements are relatively rare, about one in seven.

Ethics and Self-Interest

This gives us an extra piece of information. One reason people often give for not using randomized clinical trials is that they are unethical. That is, one should not give a patient an inferior treatment. The information obtained from the controlled studies of surgery and anesthesia that we reviewed showed that the physician does not know which way a trial will come out. This goes far toward resolving the ethical issue.

In a period when the population is tightening up its attitude about participation in experiments, in sample surveys, and generally in information-producing activities, we need also to think about self-interest. Let us focus on the medical situation.

Sometimes participation in a trial may directly help the patient. The patient may be lucky and get the preferable therapy, or the treatment may be reversible so that after the trial, patients who had the less useful treatment can be given the better one. Nevertheless, sometimes the outcome of the trial may be of little benefit to the individual or his or her relatives or friends.

In spite of this, the patient may still wish to participate in the trial. If we recognize the trial as part of a general system of trials in which patients participate only when they qualify and when we require a trial to find the better treatment, we see that the patient may benefit not from the particular trial but from the system of trials.

Findings from other trials will help the patient, or relatives or friends. We should have our eye on the pooled benefit of the whole system. The longer the patient lives, the more likely he or she is to suffer from some of the diseases that we learn about through this system of trials. The patient will then be the beneficiary of this information. If trials are not made, then the information will be slow in coming, if it comes at all. Thus the patient has a stake in the whole system, not just in a particular trial.

The inferences derived from trials apply to the populations who participate in them. To the extent that individuals decline to participate, and to the extent that their responses may differ from those of others, the treatments may not apply as well to them and people "like" them. If special groups either deliberately fail to participate or if they are barred from participation, then the trials cannot be expected to apply as well to them as to the groups who do participate. It is hard to say just what "people like me" means, and a good solution is to have volunteers from the whole appropriate population.

If participation seems to be a sacrifice, others are making similar sacrifices in aid of "my" future illnesses, and the whole system is being upgraded for "my" benefit. Thus a special sort of statistical morality and exchange needs appreciation (*9*).

Measuring benefits. In studying costs, risks, and benefits of surgery, we found that measuring benefits was our weakest point. Survival is the most-used measure of benefit. But much surgery, maybe most of it, is intended not as lifesaving but for improving quality of life. This means that we need to assess quality of life (convenience and comfort) to find out how much we are improving matters. Before we try this in social programs generally, we would do well to develop our methods in an area like surgery. My colleagues and I have been trying this out on an exploratory basis using a brief questionnaire, and we are much encouraged.

Safe surgery dilemma. Information about economics, about outcome, and agreed-on ethics cannot entirely determine social policy. What I call the safe surgery dilemma illustrates this (*10*). Consider a safe surgical operation like that for appendicitis. If physicians operated whenever they saw even slight signs and symptoms, the total lives lost might be minimized. But if they operated only when the symptoms were severe, this would minimize the total number of days spent by patients recuperating.

None of us want to be operated on needlessly, nor do we want to die because we have shown only mild symptoms. What should the policy be? To save the last life may require a million extra operations or hundreds of lifetimes of recuperation. This is a problem that society must settle, and while information and ethical considerations can help, the decision is a social one. Note that the conflict is not the usual one between society's interests and those of the individual, but is a conflict within the individual. We have here the classical mathematical difficulty that we cannot expect to maximize two functions at the same time. That is why the happy principle of "the greatest good for the greatest number" is only a slogan and rarely a useful tool.

Linkage and confidentiality. While I am on the subject of information, let me mention a further matter. A great deal of valuable information about the economy and about health is tied up in government computer systems. It is difficult to relate various kinds of information for statistical study purposes because we have become more and more concerned about privacy and confidentiality.

This leaves us with a serious question. Are we going to purchase this information all over again within a time delay in order to solve new problems, or are we going to use what we have? To use it requires that we link up information about an individual from separate statistical series. We do not need to know the person's name after the linkage has taken place, but some identification is required to make the link. Under suitable auspices, the linkage can be made and then the names erased. After that, statistical analyses can be made.

As an example of the need, associated with evaluation, we have many statistical series in the United States of exposure to various chemicals. Let us call these the input information. We also have many series concerning deaths or disablements or morbidity from various diseases. Let us call these the output information. What we do not have is many series where the exposure input data are linked to the health outcome data. A recent study at the National Center for Health Statistics (*11*) found only four series that related environmental input to health outcome, and the linkage was primarily of a geographic aggregate nature rather than on a single individual basis.

Thus if we want to clean up the environment, we need data linkage to tell us how to spend our money to reduce damage to health. We need to know where the most damage occurs and how effective expenditures would be in reducing health losses.

If we knew that one social policy would save a person-year of life for each $10,000 spent, and another policy would cost $500,000 per year of life saved, this information might well influence us in deciding how to spend the money. Lest you suppose that such extremes do not

arise, let me say that we can document wider extremes in some current lifesaving and safety policies (*12*).

What sort of strategy might we have for getting this information? We need more linkage of data about individual exposure and life histories and their relation to health outcome. We are trying to control health events that may take 20 or 30 years to develop, and we do have data of potential value in choosing such controls. The self-interest of the society might well decide that instead of starting out now to gather such data from scratch, we would do better to have some linkage and consider the various amounts of damage that different forms of exposure create, and what it might cost society to reduce untoward effects. Nevertheless, this is a political issue, and society may prefer its privacy and confidentiality to providing information that may save lives and disablements.

Research on Research

There have been several studies of basic research, or perhaps of research in general. We need many studies in this area, not to discover whether basic research yields dividends, but to find out something about the prospects for success of various kinds in research and development.

The first such study was Project Hindsight, which was carried out by the U.S. Department of Defense (*13*). Its general conclusion, which did not cheer up basic researchers, was that basic research did not contribute much to the development of weapons systems. It concluded that targeted research contributed more.

The second was the study of Comroe and Dripps (*14, 15*). This study traced some major biomedical developments to their basic research roots, and showed the essential role of basic research in inventing new therapies.

The go-no-go approach to basic research seems to be a not very helpful concept. We need basic research for new developments. The problems must be what basic research is needed, and how much is worthwhile in a given area. Can evidence be adduced which would help with the funding and educational and occupational decisions that must be made? It is one thing to say that nobody knows, but another to face the fact that someone has to decide how much money to provide and how to spend it for the public good. Although such questions are political, definite quantitative information can help us with such decisions. Let me describe what I found common to the Comroe-Dripps and the Hindsight studies in spite of their opposing conclusions.

The first finding was that major practical advances in both weaponry and biomedical therapies seemed to require not just one innovation or breakthrough, but a bundle of them, often as many as a dozen. The second was that there is a substantial period, often 20 years, between a basic science innovation and its use in weaponry or therapies. If a variety of new things have to be assembled to make a whole, it is not surprising that they might on average be somewhat aged before being used in a major innovation.

Comroe and Dripps studied the origins of the ten most important clinical advances in cardiopulmonary medicine and surgery occurring between 1945 and 1975. Of 529 key research articles leading to these advances, 41 percent "reported work that, at the time it was done, had no relation whatever to the disease that it later helped to prevent, diagnose, treat, or alleviate" (*14*, p. 12).

A report with the acronym TRACES (*16*), prepared by the Illinois Institute of Technology Research Institute for the National Science Foundation, dealt with five advances: magnetic ferrites, video tape recorder, the oral contraceptive pill, the electron microscope, and matrix isolation. By studying a longer time period than Project Hindsight, the investigators found that key events leading to these advances divided into three groups: 70 percent nonmission research, 20 percent mission-oriented research, and 10 percent development and application. The distribution of nonmission events had a mode between 20 and 30 years prior to the innovation, while mission-oriented events peaked during the decade prior to the innovation. For these case studies, time from conception to demonstration ran about 9 years. Ten years prior to the innovation, 90 percent of the nonmission-oriented research had been completed.

The Battelle Columbus Laboratories (*17*) extended this research by adjoining the heart pacemaker, hybrid grains and the Green Revolution, electrophotography, input-output economic analysis, and organophosphorous insecticides to the magnetic ferrites, video tape recorder, and the pill studied by TRACES. The average time from conception to first realization of the innovations was 19 years. This set of innovations took longer to realize than those of TRACES, and of the significant events leading to the innovations, the distribution was 34 percent nonmission, 38 percent mission-oriented, 26 percent development, and 3 percent nontechnical. Thus the distribution of key events into the categories varies depending on the choices of innovations to study and perhaps on who classifies them. It seems clear, however, that both mission- and nonmission-oriented research are important, and that the nonmission work goes on generally well in advance of the mission-oriented research, which in turn tends to precede the developmental work.

We need some additional kinds of studies that are retrospective and prospective. For example, we need to have an idea about the population of research being done, and what it emits, in addition to a method that starts with highly selected output and works back.

Once this idea has been worked over carefully so that we understand what we need to find out, we then might engage in a truly prospective study. That is, we ultimately need to develop a study based on research as it starts. The major difference between a forward-looking retrospective study and a prospective study is that we have the opportunity to gather the data we want in the prospective study. In the forward-looking retrospective study we have to make do with the data history has provided. Recollections often differ, and the older I grow, the more I distrust oral history.

Funding agencies generally, and the U.S. Congress in particular (*18*), especially desire more research of this kind. We do not know much about how to do it. Blume (*19*) says that we should not expect universal principles of scientific management, but that the comparative analysis of scientific communities might do much to help us understand the workings of science. In such studies, he says, we might find out how organizational factors, resources, and division of labor vary in their effects from one specialty to another.

Although we can scarcely instruct anyone how to do this research on scientific productivity and scientific management, we should encourage a good deal more of it and not expect much payoff soon.

Successful technological innovations. We need considerably more work in the area of research on research, both in basic science and in innovations in technology. I illustrate this for the technology side using the British study called Project Sappho.

Investigators at the Science Policy Research Unit at the University of Sussex studied the reasons for success and failure in industrial innovation (*20*). They

combined a matched-pair and a case-study approach. They chose instances in which a technological innovation had been introduced at least twice, at least once successfully and at least once unsuccessfully. Then they carefully applied the case-study method to the details of both the successful and the unsuccessful innovation. In all, they studied 29 pairs of innovations, drawn from either the chemical industry or the scientific instruments industry. They wanted to find the characteristics that separate winners from losers.

The main finding was that no one variable seems to distinguish successful from unsuccessful innovations. Beyond this, their detailed findings can be summarized as follows: Successful innovators better understand user needs; pay more attention to marketing; develop more efficiently, but not necessarily faster; make better use of outside technology and advice; have responsible individuals with greater seniority and authority (mostly the business innovator rather than the technical innovator).

Many features did not seem relevant though often mentioned in business lore: size of firm, management techniques, use of qualified scientists and engineers, timing (being first or second to market the innovation), initial familiarity with markets and technology, structure of research, in-house versus out-of-house ideas, market pressures.

Of course, the sample studied has special features. It does not discuss technologies where just one attempt at introduction succeeded or failed, and these might form the majority of cases. Thus it would be valuable to have some further studies. In examining single successes and failures, we cannot readily create the comparability that the matching in Project Sappho provided.

Scientists and engineers. In recommending research on research and development, I wish that I could say that research on scientists and research on engineers were mutually supporting efforts and that what works for one works for the other. Research at Massachusetts Institute of Technology suggests that this is not true.

Even for research scientists working in the same firm with engineers, the goals are not the same (*21*; *22*, p. 310). Their priorities are almost reversed, with scientists more oriented to the world outside the company and engineers turned more inward toward the company.

Allen (*22*) reports that, for engineers, ideas suggested by people outside the firm for solving company problems have a low success rate compared with ideas developed within the firm. Research scientists, in contrast, find that suggestions from outside the firm have a good success rate.

Allen reports another difference. At first his group felt that engineers did not read the literature while the research scientists did. Further investigation showed that a few engineers acted as technological gatekeepers. They read the literature and interacted with the rest, keeping them informed. Of course, these gatekeepers were soon promoted to management, where they no longer interacted technically with the engineers and could no longer follow the literature. Among research scientists, the tendency was for each person to keep up with an appropriate literature—each scientist acting as his or her own technological gatekeeper.

These remarks merely support my earlier point that we cannot expect our research efforts to have universal applicability for scientists and engineers.

Technological innovation and the economy. In late 1979 the American Chemical Society held a symposium on innovation and research (*23*). In June 1980 the AAAS held its Fifth Annual R & D Colloquium (*24*). Both the symposium and the colloquium considered what could be done to stimulate innovation.

Edwin Mansfield (*25*) pointed out that the economist Zvi Griliches (*26*) used data from about 900 manufacturing firms to indicate that a firm's rate of productivity increase is directly related to the amount it has spent on R & D. Nestor Terleckyi (*27*) has shown corresponding results for whole industries. Mansfield (*28*) has found that there is a direct relation between amount spent on basic research and rate of productivity increase after adjusting for total R & D expenditure.

The participants at these conferences had many suggestions for increasing the rate of innovation mainly through changes in government policy. Several speakers pointed to Germany and Japan where they felt that the cooperation between government and industry to promote industrial development was a pattern to emulate. Others encouraged more relationship on what Daniel Boorstin (*29*) calls "the fertile verge" between industry and universities. Improved patent policy would help, some say. Others suggested reductions in regulations.

I shall not try to pull together or evaluate these suggestions made by others more qualified in this area than I. William D. Carey (*30*), Executive Officer of AAAS, pointed out that creating a substantial turn-around in public policy toward innovation would be a lot to expect because innovation has a rather small constituency. To enlarge it would require cooperation among large and small industries, labor organizations, economists, professional groups, media, and elected representatives. He doubts this will happen.

He points out too that since innovations take perhaps 10 years to develop, it is hard to evaluate the effectiveness of any specific policy shift in the process of technological innovation. He sums up these complications in a quotation from a friend who says "On a clear day, you can see practically nothing."

AAAS Initiative in Science and Engineering Education

With such a complicated outlook on the government and industrial side, what more can AAAS do? A fundamental ingredient in both scientific and technological innovation is the strength of the scientists and engineers who are available. Another important component is a well-educated public who can appreciate the value of research of all kinds and recognize the need to nourish them.

In the United States we have seen an erosion in education in science and mathematics both in amount and quality. The citizen has become less well informed, as we know from the many studies by the National Assessment of Educational Progress.

It may take 10 years to develop a technological innovation, but it takes 20 years to make a citizen or a scientist or engineer. We must find methods to do this better. As Antoine de Saint-Exupéry said (*31*, p. 155), "As for the Future, your task is not to forsee, but to enable it."

For example, I have been impressed with the educational work of the Ontario Science Centre. After Dr. Tuzo Wilson introduced me to it, I encouraged our AAAS Committee on the Public Understanding of Science to review the possible use of science and technology centers and museums as a resource for strengthening science education. That committee, chaired by John Truxal, has already taken steps in that direction. This is just one step. The committee has made several recommendations to the Board of Directors designed to strengthen the AAAS effort in education.

The board, after being informed of recent studies of science education in various countries, has resolved to mount a program to help improve science and engineering education both for the citizen and for professionals. James Rutherford will advise the board on developing a program for science education.

Our first steps will be to work with our affiliated science and engineering societies to advance science and engineering education in the 1980's. We will urge the President of the United States and the Congress to address the need for excellence in science and engineering education. I hope, personally, that our Canadian members will join us in considering what cooperative steps may be taken.

Finally, for the 1982 AAAS Meeting in Washington, D.C., we have directed that a major theme be "Toward a national commitment to educational excellence in science and engineering for all Americans."

Conclusion

To get and retain the benefits of social, medical, and technological innovations, we need to evaluate their effectiveness in practice. Otherwise we will find ourselves paying for poor innovations both in dollars and in delay of introducing better ones.

Individuals have some self-interest in participating in experiments and sample surveys and allowing the linkage of information for statistical purposes, because innovations may not be suited to the nonparticipants, and we may mount expensive programs that have little payoff. Still society must decide whether it approves. Solid information helps decisions about some innovations but cannot settle some issues where different payoff evaluations lead to differing policy actions.

To help the process of innovation and to inform the funding agencies and the public, we need more research on the research process itself both in basic science and in technology.

To maintain a strong economy we require constant innovation. This requires vigorous basic science and technological R & D programs. Providing these requires a citizenry well-educated in science and technology as well as strong training for specialists in science and engineering. This education needs a great deal of improvement. The AAAS plans to join with other societies and institutions in strengthening these educational programs.

References and Notes

1. *The Holy Bible*, King James Version, Daniel I: 12-15.
2. S. Purchas, *Hakluytus Posthumus or Purchas His Pilgrimes: Contayning a History of the World in Sea Voyages and Lande Travells by Englishmen and others* (1625) (reprinted by James MacLehose and Sons, Glasgow, 1905), vol. 2, pp. 392-398.
3. S. Davidson, A. P. Meiklejohn, R. Passmore, *Human Nutrition and Dietetics* (Williams & Wilkins, Baltimore, 1959), p. 419.
4. J. Lind, *A Treatise of the Scurvy (1753)* (University Press, Edinburgh, reprinted 1953).
5. *Encyclopaedia Britannica* (Benton, Chicago, 1964), vol. 20, pp. 231-232.
   Jane Teas has kindly called my attention to C. Lloyd's "The introduction of lemon juice as a cure for scurvy" [*Bull. Hist. Med.* **35**, 133 (1961)]. Lloyd attributes the time lag of introducing lemon juice in the British Navy primarily to misleading reports from Cook's voyages and reports from Cook himself and secondarily to the many competing remedies, failure to experiment properly, and to Lind's own lack of prominence. Bucknam McPeek has shown me "Vitamin C status of submariners" by S. C. Gilman, R. J. Biersner, and R. D. Thornton (Report 933, Naval Submarine Medical Research Laboratory, Submarine Base, Groton, Conn., June 1980). The authors find lowered plasma vitamin C in submariners after patrols, and encourage vitamin C dietary supplements, closing "After 400 years, the mistakes of the past should not be repeated."
6. J. P. Gilbert, R. J. Light, F. Mosteller, in *Evaluation and Experiment: Some Critical Issues in Assessing Social Programs*, C. A. Bennett and A. A. Lumsdaine, Eds. (Academic Press, New York, 1975), chap. 2, pp. 39-193.
7. K. S. Ditman, G. G. Crawford, E. W. Forgy, H. Moskowitz, C. MacAndrew, *Am. J. Psychiatry* **124**, 160 (1967).
8. J. P. Gilbert, B. McPeek, F. Mosteller, in *Costs, Risks, and Benefits of Surgery*, J. P. Bunker, B. A. Barnes, F. Mosteller, Eds. (Oxford Univ. Press, New York, 1977), chap. 9, pp. 124-169.
9. See J. P. Gilbert, B. McPeek, and F. Mosteller [*Science* **198**, 684 (1977)] for a more detailed treatment of this issue.
10. F. Mosteller, *J. Surg. Res.* **25**, 185 (1978).
11. B. B. Cohen, J. A. Perlman, D. P. Rice, L. E. Duncan, S. C. Marcus, M. D. Kunst, P. E. Leaverton, W. F. Stewart, *Environmental Health: A Plan for Collecting and Coordinating Statistical and Epidemiologic Data* (National Center for Health Statistics, Government Printing Office, Washington, D.C., 1980).
12. N. J. Bailey, *Reducing Risks to Life: Measurement of the Benefits* (American Enterprise Institute for Public Policy Research, Washington, D.C., 1980).
13. C. W. Sherwin and R. S. Isenson, *First Interim Report on Project Hindsight* (Office of Director of Defense Research and Engineering, Washington, D.C., 30 June 1966, revised 13 October 1966).
14. J. H. Comroe, Jr., *Retrospectroscope: Insights into Medical Discovery* (Von Gehr, Menlo Park, Calif., 1977).
15. ——— and R. D. Dripps, *Science* **192**, 105 (1976).
16. *Technology in Retrospect and Critical Events in Science* (TRACES vol. 1, report to the National Science Foundation, prepared by the Illinois Institute of Technology Research Institute, 1968.
17. *Interactions of Science and Technology in the Innovative Process: Some Case Studies* (final report to the National Science Foundation, prepared by Battelle Columbus Laboratories, Columbus, Ohio, 1973).
18. R. N. Giaimo, in A. H. Teich *et al.* (*24*), pp. 29-37.
19. S. S. Blume, *Science* **207**, 48 (1980).
20. Science Policy Research Unit, University of Sussex, *Success and Failure in Industrial Innovation* (Centre for the Study of Industrial Innovation, London, 1972).
21. R. R. Ritti, *The Engineer in the Industrial Corporation* (Columbia Univ. Press, New York, 1971).
22. T. J. Allen, *Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information Within the R & D Organization* (MIT Press, Cambridge, Mass., 1977).
23. W. N. Smith and C. F. Larson, Eds., *Innovation and U.S. Research: Problems and Recommendations* (ACS Symposium Series 129, American Chemical Society, Washington, D.C., 1980).
24. A. H. Teich, G. J. Breslow, G. F. Payne, Eds., *R & D in an Inflationary Environment: Federal R & D, Industry and the Economy, Universities, Intergovernmental Science* (American Association for the Advancement of Science, Washington, D.C., 1980).
25. E. Mansfield, in A. H. Teich *et al.* (*24*), pp. 84-91).
26. Z. Griliches, in *New Developments in Productivity Measurement and Analysis*, J. Kendrick and B. Vaccara, Eds. (National Bureau of Economic Research, Chicago, 1980), pp. 419-454.
27. N. Terleckyi, *Effects of R & D on the Productivity Growth of Industries: An Exploratory Study* (National Planning Association, Washington, D.C., 1974).
28. E. Mansfield, *Am. Econ. Rev.* **70**, 863 (1980).
29. D. J. Boorstin, "The fertile verge: Creativity in the United States," an address given at the Carnegie Symposium on Creativity, the inaugural meeting of the Council of Scholars of the Library of Congress, 19-20 November 1980 (Library of Congress, Washington, 1980).
30. W. D. Carey, in W. N. Smith and C. F. Larson (*23*), chap. 23, pp. 231-234).
31. A. de Saint-Exupéry, *The Wisdom of the Sands*, translated by S. Gilbert from the French *Citadelle* (Univ. of Chicago Press, Chicago, 1979; reprint of the edition published by Harcourt, Brace, New York, 1950).

32. I thank a number of persons for help in the preparation of this manuscript: T. J. Allen, J. Bailar, W. D. Carey, J. Charette, E. E. Clark, B. J. Culliton, R. Day, R. Dersimonian, D. C. Hoaglin, P. Lavori, A. Lisbon, B. McPeek, E. Mansfield, M. Olson, K. Patterson, L. Robinson, J. Teas, D. Wolfle, C. Youtz. Preparation of this paper was aided by National Science Foundation grant NSF SOC 75-15702 to Harvard University.