The Use Of Science And Mathematics Education Indicators And Studies: A Briefing

Robert F. Boruch and Erling Boe
University of Pennsylvania

This briefing concerns the "footprints" that might be made by an array of projects sponsored by the National Science Foundation (NSF). "Footprints" here means (roughly) traces of whether and how the results of the projects were used. The object is to speculate on what uses of data or studies are worth looking for and why, and how one might discern them.

The target research of interest includes the statistical surveys sponsored by the NSF's Indicators Program, such as the Third International Study of Mathematics and Science. It includes policy-related work supported by the Studies Program, such as the examination of test and textbook contents and how these relate to the higher order thinking skills of students done by Madaus and colleagues.

This paper summarizes a longer report on the topic and capitalizes heavily on information supplied by NSF. Foundations such as the Rockefeller Foundation and agencies, the National Center for Education Statistics (NCES), and the Planning and Evaluation Service of the U.S. Department of Education have also posed questions about the value of the studies they sponsor. Their experience is also exploited here.

A major premise underlying this effort is that the data and studies on science and mathematics education produced under NSF sponsorship should be "useful." This premise is critical in that some research products are important in the long run by a variety of standards but can be regarded as useless by a variety of other standards. The premise is fundamental, but its import is debatable at the margin.

Conclusions are framed in terms of the lessons learned from contemporary social research on the use of data and policy studies. These conclusions cover essential formalities such as definition of the "use" of a data set or study, common methods of tracking use, the uncommon and underexploited methods of tracking, and planning for enhanced data or study use.

1. It is essential to define what is meant by the "use" of information and to distinguish among types of use. It is essential also to define the initial conditions and context of use.

Statistical indicators and studies of science and mathematics education may be "used" in the senses of (a) being recognized or seriously considered, (b) informing decisions, and (c) leading to actions. Making plain what is meant by data or study use is essential for program monitoring, of course, and can help to prevent egregious argument about what has been useful.

Different kinds of use must usually be discerned in different ways. The NSF-sponsored data on the U.S. rank in science or mathematics education (SME) relative to other countries have arguably influenced public debate regardless of any specific corporate or public deci-

"A major premise underlying this effort is that the data and studies on science and mathematics education produced under NSF sponsorship should be 'useful.'"

sions. The debate is traceable by examining public and professional press coverage. This debate arguably informed decisions to focus U.S. education goals on SME and were arguably followed by action—appropriation of funds for SME. The extent to which the debate informed decisions may be traceable through legislative hearings. The extent to which these decisions led to action may be discernible by observing changes in level of Federal appropriations for science and mathematics education research.

"... conventional citation counts fail to recognize influential studies that are not reported in journals."

Definition also means specifying initial conditions, context, and constraints. In the case of NSF and other foundations that sponsor the production of data, the initial conditions include institutional memory that is limited by staff rotation, formal data banks that are limited by resources, a basic science culture that puts priority on "interesting and important" rather than on "useful and important," and a foundation stress on "push the cargo out and fly on." At the individual level, the initial conditions include the roles of program staff members and their relations with aspiring principal investigators, the limits on the role of each, and the subcultures in which each operates.

2. It is easy to identify methods of tracking the uses to which statistical data and studies are put, but the methods are not commonly exploited by foundations that sponsor research.

A variety of ways have been invented to register the production and use of a data set or study. The common ones include the following:

- Counts of the publication of study results, especially publications in refereed journals and high quality books, coupled with estimates of how many scholars on average read how many articles in the relevant journals;
- Awards to a person or group, especially those made by independent professional organizations, for scholarly products generated through the study or data set;
- Popular press or media coverage of the study or its product, e.g., op-ed articles;
- Presentations in professional forums and especially in public forums in which decisions about exploiting the data are made; and
- Citation counts, notably of journal articles, books, or presentations that depend on the data set or study of interest.

Each has merit. Science journal citation counts, for instance, are an inexpensive device for learning whether certain academic audiences attend to the study. Each device, too, has shortcomings. Citation counts that focus on scholarly journals are arguably ineffective for important potential users, such as policy makers. In any event, conventional citation counts fail to recognize influential studies that are not reported in journals.

These methods of tracking the production of data and studies and their use have been identified elsewhere and are, indeed, employed to gauge an entity's performance. For instance, the U.S. General Accounting Office's (GAO) Annual Report to the Congress has in recent years included the number of studies undertaken, the number of GAO reports produced, and the incidence of congressional testimony by GAO staff.

At GAO, output indicators such as production of reports are almost inseparable from "use" indicators because most of such reports are requested by Congress and presumably used by the requestor. Nonetheless, where evidence is sufficient, the Annual Report also provides narrative information on the consequences of particular studies, e.g., reduction in fraud, waste, or abuse. Any Federal agency such as NSF, that produces studies and data that are supposed to be useful, might produce a similar report.

These simple methods are uncommon in that they are not systematically exploited by foundations or other government agencies that produce studies. NSF, for example, has no archive of publications produced by the researchers that it sponsors; it is not clear that NSF has the resources for an archive. In any event, a custom would need to be invented to assure that researchers send publications and presentations to NSF to build such an archive; a mechanism would have to be created to assure that the archive is used.

3. Statistical data and study results are woven into applied research and analysis, often in nonobvious ways. It is important to take into account imperfect recognition of a data set or study and to understand data filters and intermediary users of the information.

Low-level, persistent use of information can be important. But traces of it are often weak. Popular press reports, for example, often do not identify properly a study's sponsor, the research entity, or study's name. Refereed scholarly journals only at times properly acknowledge the specific data sets that were used in a publication.

More generally, data and studies pass through a variety of filters or, as Chris Dwyer calls them, intermediaries. The Congressional Budget Office (CBO) may reanalyze data set X, analyzed by professor Y, both sponsored by NSF. The GAO might cite the CBO's work without a reference to the NSF sponsor or the original analyst. This implies that the "reader" who is hired to track the uses of a study, or the electronic scanning strategies that are invented to track data use, must be flexible in going beyond a given user of information to the preceding one.

Identifying instances in which a data set or study is used, in literature that ranges from the popular press through policy documents and academic journals, is not easy. It requires time and competence. Those who take a temporary vow of poverty, who have both time and expertise, are a fine source of assistance in the task. They are called "graduate students" and are a natural resource for study of the matter.

An option for the future lies with the National Research and Education Network (NREN). This effort to understand how text and data can be electronically digitized and exploited easily is well underway. To the extent that NREN technology can be exploited to identify instances of "study use" or "data use," that is to the good.

4. The use of statistical data and studies is observable through direct observation and through self-report surveys. Corroboration is important. There are a variety of options.

The first obvious option is direct observation of a study's use in a meeting, by insiders or outsiders, in which

"Low-level, persistent use of information can be important. But traces of it are often weak."

indicators or studies are considered. The interested scholar may, as an independent observer, sit in on legislative or administrative meetings, to record what data or studies were considered by the meeting's participants. This tactic is often expensive, however.

An underexploited and less expensive vehicle for learning who used what data is through committees that fall in the ambit of the Federal Advisory Committee Act. Public committee meetings under the act require minutes or transcripts. Any member of the committee or any attendant of a public meeting can be a tracker of the use of data or a study. Anyone who chooses to acquire and read the minutes of the meetings is a potential expert on the use of certain studies by the committee.

A third option presumes that it is fair to ask the principal investigator (PI) of a study whether the study findings were used and by whom and when. PIs may be well informed or not. The well-informed PI should be recognized and exploited; he or she would benefit from both of these actions. The ill-informed PI might be educated by the question. The principal investigator's report may or may not be accurate. To the extent that such self-reports can be corroborated, they should be.

There is good precedent for full-blown surveys of the potential users of information, a fourth option. Recall, for instance, studies undertaken by the National Center for Education Statistics of school district staff members' knowledge about the information resources sponsored by the U.S. Department of Education. Independence of informants is an important but difficult matter.

There is less precedent for a fifth option: formal surveys of principal investigators who have received funds from a foundation such as NSF. The grant applicant who asks for more money is at times prepared to document users, e.g., the General Social Survey. But most grant recipients are not equally equipped to provide evidence about the usefulness of their work.

Doing surveys and so forth may help to provide evidence about what study or data set appears to have been useful. Prospective controlled field tests are a sixth option dedicated to understanding what could enhance usefulness of studies. Such controlled tests have been run in the mental health arena to learn, at least, that merely providing information is not enough to encourage change.

5. Peer review of research proposals to science foundations is a fundamental device for deciding whether a proposal warrants funding. More important here, the peer review process is an underexploited method of tracking the use of studies completed earlier by researchers who submit proposals.

Experts who are asked to review a research proposal can take into account the earlier performance of the researcher who submitted the proposal. The experts may consider a variety of indicators of the value of the principal investigator's earlier work. The performance indicators might include the uses to which an earlier NSF-sponsored study or data set, generated by the same or other investigators, were put.

There appears to be no uniform, formal mechanism for this kind of capitalization on external reviewers at NSF or at other foundations. Individual reviewers vary in their interest in the earlier

performance of a researcher who submits a proposal. It implies that where "use" of a study or data set is important, the data uses that are identified by a scholar who requests funding to do more data collection are important.

6. The durable civil servant is a fine vehicle for understanding what data set or study has been used.

For instance, both Murray Aborn (NSF) and Howard Rosen (Department of Labor) periodically produced "findings" for their directors, findings that could be used to argue that something happened as a consequence of the agency's investment in research. Foundation program staff who rotate through an agency arguably are not relevant to this task simply because it usually takes time for a study to be used in policy or scientific forums full-time. Charging a civil servant with responsibility to monitor data or study use is a good approach if no other options are available. With access to a phone, proposals, and final reports, this amanuensis can turn out periodic reports on the use of reports. Stake suggested that employing a group whose independence is guaranteed would be an interesting option, and this option is worth considering too. Review panels for research proposals might also be exploited productively in this effort.

To the extent that the culture of the civil service agency is changeable, engaging all career civil servants in the task of understanding which data or studies are used then seems desirable. Those who are capable of communication with both PIs and colleagues, and who wish to do so, are in a position to encourage PIs to attend to the matter. Limited resources and legitimate philosophical antagonism toward such a role for the scientist-civil servant need to be taken into account.

7. Focusing only on the use of data sets or studies is misleading. Data production methods are themselves useful products of a survey or study.

For instance, the NSF's support of the Second International Study of Mathematics and Science resulted in comparative data on mathematics achievement. The thoughtful tracker of the uses of data might reckon that the adoption of higher quality survey methods and testing methods is no less important. And indeed, there appears to have been an improvement in the international studies in that principals have agreed upon definitions, e.g., of 9-year-olds and grades, and methods of sampling that make cross-national comparisons more sensible.

Data on the use of methods may be available through self-reports, through monitoring attempts to augment or piggyback on national surveys, monitoring the adoption of survey data or methods in local surveys, and so on.

The slogan "technology transfer," though trite, is apropos. The methods of measurement of academic achievement, the methods of sampling, and so on that are a product of foundation investments are important. The adoption of these methods is important. It ought to be tracked.

There are good precedents for expecting that new methods of producing data are as important as the data set's implications. Precedents for the adoption of new data collection methods are easy to find. For example, randomized controlled tests of programs in criminal justice are now common partly because of the Minneapolis Domestic Violence Experiment. Randomized clinical trials in medicine have become frequent partly on account of the Salk Vaccine Trials.

"To the extent that the culture of the civil service agency is changeable, engaging all career civil servants in the task of understanding which data or studies are used then seems desirable."

"The research design issue is whether one ought to sponsor one massive study or sponsor several independent ones if the object is to assure that the resultant data are used."

8. To judge from empirical study, and as one might expect, certain variables are related to the use of information. The implication is that further empirical study is warranted and, more important perhaps, that one might statistically impute the use of data sets and studies rather than observe their use directly.

There is good empirical evidence that what matters in assuring that data sets or studies are used includes variables such as the potential users' access to the data or study, the quality of the data, the context and complexity of use, and the background of the potential or actual data user.

Each of these variables is in some sense observable. In the absence of any opportunity to directly observe data use, one might impute the use of data from observations on such variables and a simple statistical model that relates the outcome variable—use—to these variables. There appears to have been no published work on such an effort.

9. Policy, strategy, and systems for data use enhancement are important and warrant special study.

Sponsors of studies have helped to enhance the likelihood that a policy-relevant data set or study will be used, notably by investing funds in dissemination, e.g., Rockefeller Foundation's investment in underclass research. Sponsors have been less sensitive to assuring that the effect of this investment is discernible. The Rockefeller Foundation is an exception in that it has asked for independent review of its investment in both policy research and the dissemination of research results.

Data-sharing policy has been adopted by NSF, the National Institute of Justice, the National Heart, Lung, and Blood Institute, and other organizations that sponsor data production and research. This is remarkable relative to many other agencies and foundations. Tracking the sharers is warranted, however. None of the data-sharing agencies have a tracking system, and this invites the invention of a low-cost independent tracking system.

Data enhancement policies and systems that include piggybacking, sample augmentation, and satellite design are promising. For instance, that several agencies cooperate in trying to produce a useful product is worth recognizing. Presumably, all agencies thought the need for the data set or study was sufficiently important to collaborate in the effort to produce it. The collaboration is an easily measured phenomenon and may be taken as an indicator of expected usefulness of a study or data set.

The option of designing multiple independent studies or multiple loosely coupled studies, instead of a single massive study, deserves more attention. The research design issue is whether one ought to sponsor one massive study or sponsor several independent ones if the object is to assure that the resultant data are used. It is certainly easier to manage a big study rather than several smaller ones. But if multiple studies rather than a single study invite more uses then planning multiple studies rather than a single massive study may be productive.

10. When it is important to assure that data or studies are useful in the policy-making process, staying close to the process is crucial. Keeping distant from the policy maker is crucial, too, in the interest of credibility at least.

To the extent that the indicator/study is close to a policy-making process, the closeness can be monitored, for instance, through logs on who spoke to whom, why, and when. Telephone records, speaking records, and so on are vehicles for tracking.

Gaining the distance that is needed to assure credibility, while keeping close, is harder to do. It is not clear how to observe this.

To judge from contemporary empirical research on data use, however, credibility of the source of information that is purported to be useful is important. It is for credibility reasons that some institutions such as the National Center for Education Statistics and the Bureau of Labor Statistics separate the data production function, which ought to be more independent of politics, from the data use function, which ought to depend on the body politic. The General Accounting Office is similarly sensitive to such issues, but meets its concerns in ways that differ from those used at the statistical agencies.

The source of support for a data set or study is also important. To the extent that a sponsor such as NSF or NCES is viewed as dispassionate, the information may be regarded as credible. The public and others do at times register opinions about credibility of sources of information and of sponsors. Formal surveys of credibility of either are possible in principle, but it is not clear how to do this economically.